Is there a strong match between the construct of general proficiency in the minds of native speaker non-teachers and native speaker teachers? An exploratory study

Gautschi, Curtis

Erstveröffentlichung / Primary Publication
Magisterarbeit / M.A. thesis

Is there a strong match between the construct of general proficiency in the minds of native speaker non-teachers and native speaker teachers? An exploratory study.

by

Curtis Gautschi

Submitted for the degree of Master of Arts, Linguistics (TESOL)

Department of Culture, Media and Communication
School of Arts
University of Surrey

July 2006

Summary:

In spite of the fact that assessment criteria in general proficiency testing (GPT) is an area of extreme importance, being the expression of the underlying general proficiency construct, little has been said in the literature regarding the source of such assessment criteria. Unfortunately, even less has been done towards answering the important question of whether the application of exclusively teacher-based perceptions to GP assessment criteria is sound, given the fact that context of use and test interpretation involves the general public. This paper outlines a rater cognition study that explored the GP construct by examining unguided reactions of 2 non-teacher native speakers (NSs) with 2 NS teachers. Notable quantitative and qualitative differences were found. The findings suggest that the construct of GP as rooted in the perspectives of teachers trained in the use of rating scales has questionable utility to the actual contexts where GP test scores will be interpreted, and where the abilities measured in GP tests will actually be used.

Acknowledgements:

I would like to thank the speakers and raters who gave up some of their free time to participate in this study, Szilvia for her helpful support, and my family: Mom, Dad and especially Mary and THE KIDS. It is no exaggeration to say the blood, sweat and tears were as much yours as mine through all of this. Thank you.

Table of contents

List of Figures

List of abbreviations

ACTFL	American Council on the Teaching of Foreign Languages
ALTE	Association of Language Testers in Europe
CEFR	Common European Framework of Reference for Languages
CLA	Communicative language ability
COE	Council of Europe
COU	Context of use
ELT	English language teaching
ENS	Educated native speaker
ETS	Educational Testing Service
FCE	First Certificate in English
FL	Foreign language
FSI	Foreign Service Institute
GP	General proficiency
GPT	General proficiency testing
IELTS	International English Language Testing System
ILR	Interagency Language Roundtable
NS	Native speaker
NNS	Non-native speaker
OPI	Oral Proficiency Interview
SL	Second language
SP	Special purpose
SUPSI	University of Applied Sciences of Southern Switzerland
TLU	Target language use context
UCLES	University of Cambridge Local Examinations Syndicate
VPA	Verbal protocol analysis

Chapter 1

1.1. Introduction

To say that general language proficiency testing (GPT) is important is to state the obvious. Its importance is of course directly proportional to the importance of language ability as a key to personal mobility and opportunity. The development of proper instruments for measuring this general language ability, that are meaningful to interested parties, is naturally the task of the general proficiency testing industry. Fulcher (2003: 116) identifies four main aspects of testing: the test construct, test tasks, expected testee responses, and how responses are scored or rated. This study will attempt to explore GPT assessment criteria and how they interface with these aspects of testing. Specifically this will lead us to the important question of whether what underpins current GPT matches the perceptions of stakeholders. It is suggested that since laypeople represent a part of the intended context of use and score interpretation for GPT, the perceptions of teachers, upon which currently operationalised GP assessment criteria are defined, must be compared with those of laypeople. This will be done through a rater cognition study, where both quantitative and qualitative findings will be analysed. A match between the perceptions of those found in intended contexts of test use and the construct embedded within assessment criteria is crucial to test construct validity. Our study will then be useful towards supporting or not supporting current models and theories of general proficiency.

A review of pertinent literature will be followed by a description of this research project and the data gathering process. An analysis of our findings, as guided by our research questions, will lead to some concluding remarks regarding this study and suggestions for further research.

Chapter 2

2.1. Authenticity

In testing, authenticity can refer both to closeness of test features in terms of content, tasks and skills to corresponding areas found in the real situations, or to the match between the skills being measured and the type of test instrument being used (Davies et al. 1999: 13). The development of proper test design and the selection of test tasks are therefore key features of the testing enterprise. Bachman (1990: 244-247) states that test content and tasks are important in that they enable the proper interpretation of test scores. This is based on the view that the closer the test elicitation instrument approaches the target language use context (TLU), the more meaningful inferences will be regarding test performance (Bachman & Palmer 1996: 23), assuming a thorough understanding of the TLU (Weir 1990: 25). Generalisability of test results is clearly compromised by using inauthentic materials or test methods (Spolsky 1985: 31).

While authenticity is also important regarding teaching and classroom activities (O'Malley & Valdez Pierce 1996: 59), which are obviously key features of the larger context of teaching and testing, there is a noticeable lack of concern for authenticity in terms of assessment criteria. The literature instead appears to be primarily concerned with authenticity in testing in terms of test content and tasks. McNamara (1996: 38) states that SL performance testing rationale is based on the focus on real life tasks. Salaberry (2000: 297-299), in noting the lack of substantial changes from 1986 to 1999 in the revision of American Council on the Teaching of Foreign Languages (ACTFL) procedures for eliciting speech samples in the Oral Proficiency Interview (OPI), notes this almost exclusive focus on task and test context. Lewkowicz (1997:168) notes the narrow perspective of Bachman & Palmer (1996)'s definition of authenticity in terms of correspondence with the TLU task, suggesting that the concept of authenticity has not been adequately studied. Unfortunately Lewkowicz makes no call to investigate authenticity in terms of assessment criteria. While there is concern over this issue in the literature, it does not come from GPT but

rather from the Special Purpose (SP) language testing field: "writers on performance assessment in general vocational contexts have focused almost exclusively on the stimulus and the process, ignoring the simulation of real-world criteria" (Jacoby & McNamara 1999: 214). It should be remembered that this overwhelming concern for authenticity in terms of test tasks was partly in reaction to the practice of measuring aspects of speaking ability such as pronunciation and syntax, independent of real communication acts (Savignon 1972: 11-12 in McNamara 1996: 32). The concept however must now be expanded to include authenticity in all areas of testing, including criteria of assessment.

2.2. Assessment criteria

Weir (1990:11) states that "If inauthentic tasks are included in tests of communicative language ability there is a real danger that the method employed could interfere with the measurement of the construct we are interested in." It is suggested here however, that the same is true given inauthentic assessment criteria. Indeed, Clark (1972: 128 in McNamara 1996: 32) states that the basis of proficiency assessment is to be found in the assessment criteria as described in the rating scales, not only in the task or the test format. McNamara (1996: 32) further stresses that assessment criteria encode the very construct of the test itself that in turn necessarily reflects a specific view of language. This clearly suggests the tremendous importance of developing authentic assessment criteria.

In their seminal work over 25 years ago, Canale and Swain (1980: 25) wrote: "Little serious attention has been devoted to criteria for evaluation and levels of achievement/proficiency with respect to a given theory of communicative competence". While theories of communicative competence or language ability (CLA) such as the immensely influential work of Bachman (1990) have been developed since, based notably on the work of Canale and Swain (1980) and others, a careful look at the present state of affairs suggests that considerable energy has gone into areas quite different from those suggested by Canale and Swain, as we shall see below.

2.3. Source of assessment criteria

In our discussion of the source of assessment criteria it is necessary to look at the key relationship between the construct of proficiency or general language ability and its outward manifestation a performance. The reader will note that here and throughout, proficiency, ability and competence are used interchangeably, as a distinction between them is difficult to maintain (Davies 1999: 153).

Language proficiency may be understood in terms of a pragmatic description – "person x is able to do y", or in terms of a theoretical construct – "person x has ability y" (Bachman 1990: 254). Naturally, "being able to do y" implies a communicative act within a context, and requires a judgement: "x is in fact able to do y". These descriptions suggest that pragmatic "proficiency" is analogous to "performance" and that proficiency and performance are necessarily linked, the former referring to what is within an individual, the latter when it is manifest in a context with others (Fulcher 2003: 20). Weir (1990: 7) simply states that it is difficult to maintain any difference between the two, noting that nothing can be known at all about proficiency nor can it be measured, except through performance (see also Bachman 1990: 256). Nevertheless, the relatedness and the location of the two, one within an individual, the other in interaction, is problematic as Bachman (1990: 308) notes: "the distinction between language ability and the performance of that ability has been at the same time a central axiom and a dilemma for language testing".

Let us examine this dilemma with reference to GPT and specific purposes testing (SPT) briefly. In GPT a set of test tasks is selected to tap into the hypothesised proficiency, which in the context of the test will stimulate a performance. This performance implies the presence of proficiency, which is then generalised to predict performances in other contexts beyond the test. Indeed, the whole objective of communicative language testing is to make inferences from a set of performances within the test, to non-test contexts of use, that are, in principle, not observable in advance (McNamara 2003: 467). Douglas (2001: 172) claims this type of general

purpose testing is necessarily theoretical since the future context of use (COU) is unspecifiable. At the other end of the continuum, we find SPT, where the context of use is known and specifiable (Douglas 1997: 111). SPT has the potential to be empirically-based in terms of construct, tasks, performances and attendant judgements of those performances, since it is possible to examine the future COU and create the test accordingly. On the other hand GPT lacks this possibility since future contexts of use are unspecifiable, and therefore GPT necessarily remains theoretical. It is suggested that this almost fatalist perspective has in part led to the equally theoretical nature of rating scale development and assessment criteria selection in the traditions of operationalised GP tests in use today.

To say that the future COU is not specifiable is rather misleading. It implies that nothing may be known beforehand regarding the situations where learners will use language after the test. This is quite untrue, since all future COUs are real contexts, and to some degree can be known. The problem is rather that the possible COUs and related discourse domains are too numerous or too broad for easy examination and classification. In any case, this dilemma will not go away on its own. If GP is too complex or broad to be specified, the construct of GP is non-falsifiable and therefore a questionable base for any operationalised rating scale framework. A discourse domain is identified by its content, and defined by its context. If these cannot be meaningfully examined by definition, this strongly diminishes utility of test scores since inferences regarding future performances are equally meaningless and beyond verification.

The above argument does not however take into consideration the fuller context of any performance, either in a test or in the real world. We may recall that performance is contextually bound to an act of communication.

> Language cannot be meaningful if it is devoid of context (linguistic, discoursal and sociocultural). (Weir 1990: 11)

> Established understandings of the construct of language proficiency have been challenged on the grounds of a variety of more socially oriented conceptions of language use. (McNamara 2001: 333-334)

Accepting that communication is thus socially and contextually defined in a performance, the quality of that performance must also be contextually defined, specifically in the minds of the other participants in context. This is in essence the premise behind Douglas (2001)'s study, where the construct of perceived proficiency is explored as a means to developing assessment criteria that are found in the COU or the target language use situation.

It may be objected that Douglas (2001: 171) is addressing specifically SP contexts when arguing in favour of deriving assessment criteria from the target language use context. While true, the central point that the perception of proficiency is located in the minds of those within the TLU and that the construct of proficiency is embedded within the target language use context, is nevertheless crucial to a proper understanding of assessment, including GPT.

2.4. Source of assessment criteria in current GP frameworks

In this section we will look at currently operationalised testing frameworks or tests that are based on the general proficiency construct.

2.4.1. American Council on the Teaching of Foreign Languages

The ACTFL Speaking Guidelines and Oral Proficiency Interview (OPI) have been extensively criticised in the literature, most notably by Lantolf & Frawley (1985, 1989) and Fulcher (1987, 1996a, 2003). These criticisms centre on the lack of proper definition of the test construct as embedded in the rating scales. Fulcher (2003: 16) states that even the 1999 revisions of the Guidelines still amount to band descriptions of proficiency defined *a priori* with little empirical support (see also Salaberry 2000). While Lantolf & Frawley (1988: 187) go further, claiming that the entire psychometric posture of the Guidelines, based on the assumption of the "scalability

of human behaviour" is flawed, our present inquiry will not allow us to discuss this important area. For our purposes, let us accept psychometrics, and note that the Guidelines and the OPI rank performances along a hypothetical continuum broken up into arbitrary proficiency bands ranging from zero to total mastery (Stansfield 1992: 1-2) as represented in the hypothetical educated native speaker (ENS) (see Lowe and Liskin-Garsparro 1987: 4). Historically, this construct finds its origins in its predecessors, the post-WWII US Foreign Service Institute (FSI) OPI and the Interagency Language Roundtable (ILR) Rating Scales that standardised the rating process in 1968 (see Stansfield 1992 and Fulcher 2003). Together with the Educational Testing Service (ETS), the ACTFL and the ILR produced the Guidelines as they are now used today.

It is important to stress that in spite of the concerns mentioned above, the ACTFL group represents a significant and important breakthrough in speaking assessment. McNamara (1996: 1) places the 1950s FSI test in contrast to the discrete-point testing tradition, thus representing a more authentic and practical approach to language testing, focusing on production skills where speaking ability would be measured in real-life situations. Nevertheless, the ENS model is troubling since there are many possible ENS norms (Hill 1997: 275). Further, many NSs themselves do not reach the highest band. Placing NNS performances and inferred proficiency along an arbitrary continuum that is difficult to maintain empirically seriously strains the validity of this approach. Significantly however, this construct has made a tremendous impact on the majority of tests and rating scales used worldwide (Fulcher 2003: 171).

To summarise, we can see that the rating scales of the ACTFL, and its related test construct, are not based on any investigation of the perceptions of those found in any target language use context, but on an *a priori* set of bands entirely based on the construct of the ideal ENS (Stansfield 1992: 1-2).

There have been attempts to apply different approaches to measuring speaking ability. Fulcher's own research into empirically-based rating scale development

(1987, 1996a, 2003) is offered in opposition to the *a priori* ENS model of the ACTFL, focusing on how learners actually behave rather than how learners ought to behave. Noticing inconsistency in the language of the ACTFL rating scale descriptors with respect to fluency markers, Fulcher's analysis (1996a) examines the effect of hesitation, back-channelling and other areas, on the perception of fluency in raters. Deemed bulky and impractical (CEFR 2001: 212), this approach to rating scale development has not been taken on. In any case, it is interesting to note that Fulcher's work, while focussing on real production of language or language sampling, uses a "teacher-rater" as his reference point, rather than what perceptions real NNSs' utterances create in the minds of people in the real world. Thus, the real interface of real social interaction in a real context is not consulted, neither in the *a priori* ACTFL nor the *a posteriori* Fulcher model.

2.4.2. Common European Framework of Reference for Languages (CEFR)

The CEFR, representing a European form of GPT, is largely based on the work of North (2000). While North's project is clearly empirical, the nature of its empirical qualities is striking. While space will not permit a thorough discussion here, the CEFR may be described as a condensation of 1679 assessment descriptors collected from a vast array of pre-existing source scales including ALTE, FSI, ILR, ACTFL, IELTS (North 2000: 181-184) into a workable and manageable framework of descriptors organised with one overriding criteria, that they make sense to teachers (selected references: North 2000: 171, 181, 185, 186, 187-189). North (North and Schneider 1998: 217) had previously noted that there was a lack of empirical basis to the selection of assessment criteria, and yet their project does not attempt to investigate empirically what assessment criteria may exist beyond the teacher perspective. Rather, the CEFR project simply uses empirical tools to substitute "intuitive and haphazard" copying from old scales into new scales, with "rational and well-founded" copying from old scales into new scales. The possibility that the old scales may be flawed is completely ignored as teacher utility is viewed as the ultimate criteria (North 2004). It may be said then that it represents a distillation of the teacher perspective of language ability.

Since the teacher perspective underpins the CEFR, its construct is necessarily dominated by teacher schemata. Such schemata would include reactions to performances and concern for learner progress (Rea-Dickins 2004: 249). Erdosy (2000: 106-115), exploring the relationship between background factors in the setting up of rating criteria in experienced raters, found that all inferred proficiency from performances was based on the understanding of a learning trajectory similar to the raters' own teaching experiences. If true that the CEFR embodies the teacher experience, and it is also true that the GPT construct does not and cannot reside in the limited context of the classroom alone, CEFR's value as a tool for informing GP must be questioned. Indeed, since the CEFR is teacher-based, and the teacher context is the classroom where students are brought along the learning continuum from zero to the maximum level possible using a classroom curriculum, it follows that CEFR based tests will likely have more in common with progress tests than proficiency tests. In fact, North (1997: 423) admits that the purpose of the CEFR is in part to monitor the progress of learners along the continuum of language learning.

It may be said that the CEFR is merely a coordinating tool that facilitates mapping tests into a network useful to practitioners. This is in fact North's (2004) view:

> A key idea always present in the development of the CEFR was to use the descriptor scales to profile the content of courses, assessments and examinations. These can then be related to each other through their CEFR profile without making direct comparisons between them or claiming that one is an exact equivalent of the other.

The Council of Europe (COE) (2001: 178) however goes further, claiming “the Framework can be used: 1) for the specification of the content of tests and examinations; 2) for stating the criteria to determine the attainment of a learning objective”. Thus, North's (2004) claim that the CEFR is not prescriptive is in fact contradicted by the COE itself.

North (2004) writes that "the aim of the CEFR is to empower and to facilitate, not to prescribe or control" and later "it doesn't try to define what should be taught (content specifications), let alone state how it should be taught (methodology)." Distinguishing methodology and course content specifications from exams and criteria for learning objectives is pointless however, as the latter have a direct impact on the former. A great number of new course books currently on the market are specifically designed to meet with the CEFR criteria (Cambridge University Press: 2006 ELT Catalogue, Oxford University Press 2006 ELT Catalogue). As well, the highly influential University of Cambridge Local Examinations Syndicate (UCLES) integrates its GP examinations to the CEFR (Saville 2003: 57). If these points are any indication, the CEFR has already become operational. The CEFR is a dominant force in testing, in "dictating the construct in assessment projects throughout Europe" (McNamara 2003: 471). The Framework is then fast becoming the *de facto* curriculum and test, to borrow Shohamy's (1997) turn of phrase.

The CEFR has not gone without criticism (Fulcher 2004, Weir 2005b), mostly on the grounds of it not being descriptive enough. Below are some areas of concern in Weir's view:

- there is a lack of consideration of variation resulting from context differences and resulting performance
- the scales are premised on an incomplete and unevenly applied range of context variables/performance conditions (context validity)
- little account is taken of the nature of cognitive processing at different levels of ability (theory-based validity)
- test tasks are seldom related to the quality of actual performance expected to complete them (scoring validity)
- the wording for some of the descriptors is not consistent or transparent enough in places for the development of tests. (Weir 2005b: 281-282)

While these assessment criteria validity problems hinder the CEFR's ability to enable its test comparability function across languages and levels in the European context, which appears to be Weir's main concern (Weir 2005b: 281), it is remarkable that no explicit criticism is given regarding the provenance of the band descriptors themselves. A comparability function is quite useless, or at worst misleading, if the

framework for comparability is based on a construct thoroughly defined by teacher utility and experience rather than representing the perspectives of those in the target language use context. North (1997: 423) clearly states that the intended COU of the CEFR is the classroom. If this construct is applied to areas outside its intended domain, any information, including test scores, derived from it may be inaccurate or misleading.

It is small wonder to find that the CEFR, which is based on the construct of progress along a learning continuum, matches the teaching schemata of bringing learners along the same continuum (North 2004). This also implies however that the CEFR is rooted in a second "ideal" in addition to the ideal ENS as embedded in the importation of ACTFL family descriptors. The second "ideal" is the "ideal SL learner", who will attain abilities as described by the CEFR - "the Framework also defines levels of proficiency which allow learners' progress to be measured at each stage of learning and on a life-long basis." (COE 2001: 1). Naturally, SLA research is the source of empirical support for this assumption of typical pathways to learning, and yet, ironically, North (2004) himself states "these levels are not the product of acquisitional hierarchies from second language acquisition (SLA) research. Unfortunately SLA research has so far only produced a partial, contradictory glimpse of what an acquisitional hierarchy might look like." This lack of support from SLA is paradoxically cited to defend the teacher-based approach, in spite of the fact that such support is crucial to its validity.

2.5. Implications

Stakeholder utility is a prime consideration in testing. Indeed in any testing there are "*a priori* obligations to stakeholders" (Hamp-Lyons 1997: 324-325) to provide useful and accurate information about testees. It is crucial then to understand what occurs when stakeholders interpret test scores. The following is offered as a model of the stakeholder – test score – testee interface (see Figure 1).

The act of interpreting a test score is in fact analogous to making inferences about a testee's properties. The hidden properties of the testee are expressed in the performance within the context of the test, resulting in a judgement of the performance within that context by making use of the test's assessment criteria. The act of interpreting the test score is the same process but in reverse, where the score will attain meaning with reference to the tasks that commonly occur in the TLU, as well as the perceptions of how such tasks are performed within the same TLU. These are labelled *TLU context* and *TLU assessment criteria* respectively in the diagram. The score must be interpreted within the stakeholder's context, or more to the point, inferences will be loaded using the acts of communication and judgements/assessments that occur within the TLU as their reference point. In this model, the accuracy of any inferences of testee ability from a test score will be directly proportional to the degree that the testing context (content, tasks, environment, participants) and the assessment criteria match what is found in the TLU. Both are necessary conditions to accurate score interpretations. This implies then that too much interest in test content and tasks at the expense of assessment criteria will not lead to better tests with meaningful scores.

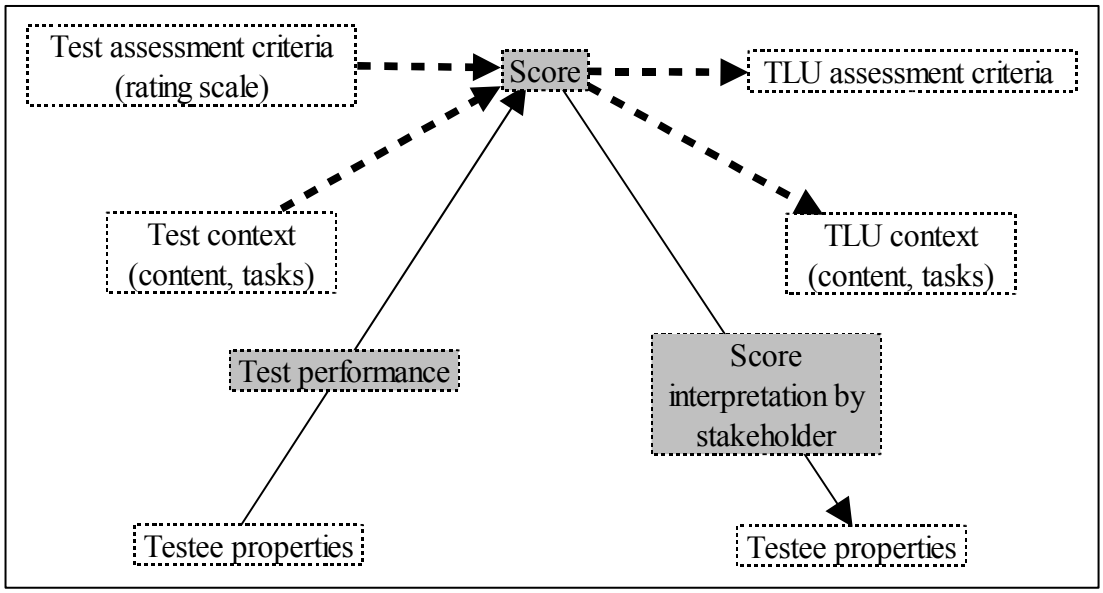


Figure 1. Model of the stakeholder – test score – testee interface

To summarise, the descriptions of expected language output and the evaluations of that output as described in frameworks such as the CEFR or the ACTFL Guidelines, and operationalised in examinations that use them, have been developed without consulting contexts of language use that learners will find themselves in following language learning and examination. Typically, they are in fact intuition-based, non-empirical and rooted in the teacher perspective (North 2000, 2004. McNamara 2003: 468). This has serious implications regarding the utility of and interpretability of test scores. If the test construct, as operationalised in the rating scales used in CEFRL informed tests, is based on a condensed teacher perspective rather than the TLU perspective, this is a construct validity problem (Bachman 1990: 242, see also Lantolf and Frawley 1988: 182, 186 for parallel arguments regarding the ACTFL), making accurate inferences regarding testee abilities difficult. If assessment criteria are not derived from the perspectives of those found within the intended context of score interpretation, they cannot be considered authentic. Finally, if inaccurate inferences result from stakeholders being unable to interpret scores within their context of interest, this represents a consequential validity problem, or plainly stated, an ethical problem, as the tests are inappropriate for stakeholders (Bachman 1990: 279, Bachman 2005, Mathew 2004: 123, McNamara 2003: 470).

2.6. Rationale for the present study

Naturally the above criticism is moot if in fact it can be empirically shown that the assessment criteria and categories of assessment as used in CEFR and ACTFL based examinations correspond significantly to those of the implied COU beyond the classroom. In other words, it may be possible to demonstrate that the teacher-schemata/classroom context construct of general speaking proficiency reasonably corresponds to the construct of proficiency as it exists in the minds of non-teachers outside classroom contexts, albeit *post hoc*. This is however practically unknown territory. The following chapter outlines our attempt to compare the perspectives of teachers and non-teachers with a view to understanding the construct of general speaking proficiency.

Chapter 3

3.1. Rationale continued: raters, inter-rater reliability and rater training

In the previous chapter, it was suggested that GP test-score utility is dependent in part on a match between the assessment criteria and categories of assessment used in GP tests and those found naturally in the target language use context. A mismatch between them undermines the possibility of correct test score interpretation. While GP testing has defined its criteria based on theoretical models of language ability (Canale and Swain 1980, Bachman 1990) and operationalised them in frameworks or tests (ACTFL, CEFR, UCLES), TLUs outside of the classroom or teacher experience have not been adequately explored, except perhaps in the SP testing tradition. The difficult task is then to explore assessment criteria in general purpose TLUs and to define them for comparison with current GP models. Before attempting this, it is necessary to consider some studies that examine raters and rater behaviour. This is crucial towards finding the right approach to defining assessment criteria and identifying the cognitive processes that underpin them in non-teacher NSs for proper comparison with those of teachers.

Once a theory of language or language ability is decided on and a rating scale is made to measure and locate performances within bands or levels, raters are needed to put the scale into practice. Clearly, scales must be applied consistently in order to provide comprehensible information about test scores to stakeholders. Errors can result from the inconsistent application of rating scales. One possible source of error is when raters attend to elements of testee discourse that are not included in the rating scales (construct-irrelevant variance – see Messick 1989, Orr 2002, Fulcher 2003, Hubbard et al. 2006). To avoid this problem, raters receive training to ensure only construct/rating-scale relevant information is contributing to test scores.

It has been said that without reliable application of rating criteria, no meaningful understanding of test scores can result (Fulcher 1987: 291). If true that "reliability is a necessary condition for validity" but "validity is the most important quality of test

use" (Bachman 1990: 289), one would expect considerable investment on the part of major testing traditions into both areas. It is certain that great effort and expense has gone towards the application of rating scales (Ffrench 2003: 8-9, Saville 2003: 108) it appears however that little has gone into the source of the rating scales themselves. For example, Saville (2003: 87) points out that UCLES strives to operate with a construct of proficiency that reflects an "underlying model of language ability and the linguistic processes used in the target-use contexts". There is however a mismatch between this, and what UCLES test development is apparently based on: COE ALTE Can-Do scales, the CEFR, and development in theoretical models of CLA (Saville 2003: 78). There is no mention of studies into TLU defined constructs of proficiency. As we shall see, this present study will demonstrate the need for greater effort in this direction.

It would be useful at this point to look at a few studies that have investigated this important area of raters and the application of rating scales. While this is only a short selection of such studies, it is enough to give us a general view of the direction this important research is going at present.

A study by Hill (1997) investigated and compared Australian and Indonesian raters on a test of English as an International Language. Categories of assessment were explicitly given to all raters: overall impression, content, vocabulary, coherence & cohesion and control of linguistic features. These categories were internalised in a half day of training in the use of the rating criteria. Hill justified the use of training by stating that it is a universal practice necessary to reliability. Her study found no significant difference between NNS Indonesian and NS Australian raters' ability to use the rating scales.

A study that is frequently mentioned in the literature (Fulcher 2003, Erdody 2004, Douglas 1997b) is Barnwell's (1989) inter-rater reliability study into the use of ACTFL scales. His study compared the reactions of 14 un-trained NSs of Spanish ("naive" raters) to 1 NNS rater trained in the use of ACTFL scales. Language samples from 4 testees were elicited using an OPI, and both groups used the ACTFL

scales for assessment. Barnwell found that performances were ranked the same for all raters but with different band assessments (In Fulcher 2003: 143) The validity of conclusions from this study has been criticised on the basis that the ACTFL guidelines were translated into Spanish, and that only one trained rater was used in the study (Hill 1997: 279).

Meiron & Schick (2000) studied the progress of a group of Egyptian EFL teachers after an 11-week training course in California. While noting quantitatively measurable gains in oral skills after the training course, they also found that qualitatively different performances in an Oral Proficiency Test often resulted in quantitatively similar scores by raters in the use of the predetermined 5-category rating scale. They also found that rater background had an influence on the rating process, as unexpectedly revealed in their data. They concluded their study by stating "that there is a very real need for more studies focusing on raters, including background and training, the features of discourse they attend to, and the thought processes they employ" (Meiron & Schick 2000: 170).

Orr (2002), in part responding to Mieron & Schick's call, analysed verbal protocols produced by raters in the process of evaluating FCE speaking exams using standard FCE speaking assessment criteria. He noted a tension between raters attending to criterion-relevant and non-criterion relevant information, and argued that this discrepancy suggests a need for more in-depth rater training. He also tentatively suggested this might imply the inappropriateness of the current rating scales themselves (Orr 2002: 153).

What Meiron & Schick suggested above, may of course be done in one of two ways. One may choose to give raters a set of rating scales and assessment categories with predetermined weightings *a priori*, vary rater properties such as background, L1, teaching experience, amount of socialisation/training in the use of the scales in question and so on, and analyse what it is that raters attend to, and what score variations occur. This however leads to rather uninteresting conclusions, as it presupposes the validity of the assessment framework itself. In other words this

approach may amount to an investigation into how raters of various types react to an arbitrarily defined rating system based on an assumed ENS ideal. Unfortunately this is the pattern that emerges from the vast majority of studies, as the above examples demonstrate.

It is suggested that the above approach offers no insight into the construct of proficiency as it may exist naturally in a target language use context. In giving categories of assessment which encode a construct, the construct to be measured is being forced or loaded into the minds of participating raters. It is extremely difficult to be aware of any potential variance in the construct present naturally in the minds of raters in such cases. The second approach, which is suggested here, would be to not assign assessment categories and criteria *a priori*, but to allow the participants themselves, representative of a real TLU, to define the construct for us. This would be done by examining their thought processes and how they react to discourse. After this is accomplished, the categories and criteria of assessment may be defined *a posteriori*.

Jacoby & MacNamara (1999), again from the SP perspective, have come to this point themselves. Their study investigated the apparent mismatch between the results of an Occupational English Test and the opinions of medical experts on the criterion of success in a future bridging course. The results showed that, in using the OET rating criteria, medical professionals rated testees similarly to OET raters. Since the test and rating framework did not allow for a significant difference in scores, as was expected given the complaints from the medical professionals themselves, it was clear that the test and its assessment criteria were the problem, as they were evidently not tapping into the needs of stakeholders. Obviously, future studies must clearly work toward capturing the construct as it exists in TLUs, and then develop a testing and assessment framework to match.

3.2. Methodology

As shown above, there has been recognition of the need to investigate real contexts towards understanding the construct of proficiency in SP testing. There is however a distinct lack of research into the exploration of the GP construct as might be located in the perceptions of laypeople. It is this lack that this present study will attempt to remedy.

Since we are interested in the perceptions of laypeople and trained teachers and it is necessary to avoid the predetermining of assessment categories, an appropriate methodology must be chosen for its ability to tap into the cognitive processes of those involved, specifically towards comparing those of teachers and non-teachers. A qualitative technique is best suited to our purposes here, as it can provide insights into processes involved in performing tasks (Banerjee and Luoma 1997: 276). Specifically, a verbal protocol analysis (VPA) is the best method for our purposes, whereby it is possible to make inferences about cognitive processes (Green 1998: 1) and schemata that influence the perceptions of listeners.

3.2.1. Defining the target language use context

For the purpose of this study, the perspective of NS teachers, representing the construct as it is embedded in current models of GPT, will be compared with those of NS non-teachers, representing the target language use context of general proficiency. Of course, this is a gross oversimplification. The TLU of the GP construct is vast, a full examination would require research quite beyond the scope of this short paper. Further, not all future co-participants in communicative acts would necessarily be NSs. In the case of Switzerland, some "of the larger Swiss companies, for example banking giant UBS and telecoms (sic) operator Swisscom, now use English as the official language in many of their departments" (Foulkes 2003). Thus GP test takers may find themselves later in TLUs using English, but not with NSs at all. In any case, it should be evident that non-teacher NSs are more likely to be found in post-test TLUs than EFL teachers and therefore represent at least a part of the implied

TLU that GP testing ought to be tapping into. Our study will therefore, as an initial step, compare the NS teacher (trained rater and non-trained) with the non-trained NS non-teacher. Now we turn to the description of the research design and methodology of our study.

3.2.2. Research Questions

Our study begins with the main research question:

> Is there a strong match between the construct of GP in the minds of NS teachers and NS non-teachers?

As discussed above, this should not be investigated by simply asking teachers and non-teachers to rate performances using a predetermined set of criteria. Rather the approach used here will be the same as in Brown et al. (2005) where rater cognition is understood to be the foundation of any perception of the proficiency construct and therefore the source of rating criteria. We will begin our inquiry by exploring how people think naturally about language performances and how they form ideas of language ability. Only then can meaningful criteria of assessment be defined. Our initial research question then leads to 2 related areas:

> Is there is a strong quantitative match between the general cognitive processes of NS non-teachers and NS teachers when listening to NNSs?

Basically, this refers to what people do with what they listen to. Given the same speech samples, do NS non-teachers and NS teachers process them in a sufficiently similar manner? Of course people's thoughts may turn to many things while listening to speech. These may include making inferences about the speaker, attending to particular salient features of what the speaker says or the way it is said, or the effect speech has on the listener. Clearly, assessment criteria that have no meaning or do not occur naturally in particular contexts, if applied in those contexts will only cause confusion to stakeholders. If, as discussed above, GP testing criteria

result from the trained-teacher perspective or in other words how they process language input, and the same input is processed differently by NS non-teachers, resulting test scores from the GP test cannot be said to be meaningful to NS non-teachers.

> Is there is a strong qualitative match between the cognitive processes of NS non-teachers and NS teachers when listening to NNSs?

A qualitative match is also required for the utility of any rating scales that might be developed based on rater cognition. It may be possible that both NS non-teachers and NS teachers use sufficiently similar thought categories when listening to NNSs, but thoughts within those categories are radically different. In other words, different listeners may focus on the same aspects in the same degree but process them differently. For example, two listeners might focus exclusively on pronunciation and grammar but because of other factors have qualitatively different opinions. This final question is key to the validity of the weighting of assessment criteria. It may be that certain aspects have priority over others for some listener types. Quantities alone will not offer insights into this important area.

If a strong match is lacking in these important areas, a match between the construct of GP in the minds of NS non-teachers and NS trained-teachers cannot be said to exist. It would follow that the utility, appropriateness and validity of GP testing that is based on the teacher perspective are equally lacking. A massive reworking of the present approach to GP testing would then be required. If on the other hand, a strong match is found in the way NS non-teachers and NS trained-teachers process and react to language input, then tests and assessment criteria based on the trained-teacher perspective can be considered useful to non-teacher NSs, towards accurate interpretation of test scores. This study would, as a result, contribute to the current body of validation studies found in the literature regarding current models of GP testing.

3.2.3. Data Collection

In order to investigate how teachers and non-teachers mentally process the speech of NNSs, without imposing categories *a priori*, the following Verbal Protocol Analysis (VPA) was set up. This was done in 2 stages. In Stage 1, speech samples were gathered from 3 NNS who were interviewed and recorded. These represent the primary data used to elicit NS reactions. In Stage 2, 4 NSs listened to the speech samples, and their reactions were recorded. These verbalisations represent the secondary data that through analysis allow the comparison of what is attended to in speech.

3.2.3.1. Stage 1

The participants in Stage 1 were all NNSs with different ability levels. In using CEFR criteria, the researcher and Listener 4 rated Speaker 1, 2 and 3 (S1, S2 and S3) at B2, A2 and C1 levels. Some personal data are provided in the chart below

Speaker	*Age*	*M/F*	*Studied abroad*	*Level (CEFR)*	*L1*
S1	45	F	Y	B2	Italian
S2	20	M	N	A2	Serbo-Croat
S3	42	M	N	C1	Italian

Figure 2. Speaker characteristics

A simple elicitation instrument was selected for our primary data. While many different forms had been considered, a simple 3-part interview was finally chosen, consisting of:

- a warm-up, with some general personal-information questions
- a picture task, where the speaker interacted with a set of photographs
- a short topic discussion, where the speaker and interviewer interacted together (see Appendix A)

This format is partly based on the oral proficiency exam currently in use at the University of Applied Sciences of Southern Switzerland (SUPSI), and loosely based on the "picture task" and "long turn" sections of the UCLES First Certificate in English (FCE) exam (UCLES 2001). This was chosen for 2 reasons: 1) as mentioned above, we are interested in comparing the perspectives of teachers and laypeople when listening to speech samples elicited by a testing instrument that is currently operationalised 2) a more complex elicitation instrument might prove difficult for non-teachers to follow, hindering the data gathering process in Stage 2.

To lower the risk of listener fatigue, only parts 2 and 3 of each interview in Stage 1 (picture task and topic) were recorded for use in Stage 2. Stage 1 interview tapescripts are provided in Appendix B.

3.2.3.2. Stage 2

The 4 NS "raters", although here called "Listeners" (notation used: Li1, Li2, Li3 and Li4), listened to the speech samples gathered in Stage 1. The listeners were explicitly told to not "evaluate" but to simply verbalise any thought that came to mind as they listened to the NNSs. There was no mediation at all on the part of the researcher, except occasional reminders to keep speaking. These verbalisations were recorded, transcribed and analysed. The 4 listeners chosen were all NSs, 2 teachers, and 2 non-teachers. One of the teachers (Li4) is a trained and experienced UCLES oral examiner, the other (Li3) while an experienced EFL teacher, has no oral examining training or experience. One of the non-teachers (Li2) lives in a non-English speaking environment (Switzerland) and speaks several foreign languages, the other (Li1) in an English speaking country (Canada) and speaks only English. Although we are interested primarily in comparing the perspective of the layperson living in an English environment (Li1) with that of the trained-teacher (Li4), 2 additional participants with 2 other experiential backgrounds, were chosen for comparison purposes.

				Living in	FLs studied/ spoken
Listener 1 (Li1)	NS	non-teacher		English speaking environment	0
Listener 2 (Li2)	NS	non-teacher		Non-English speaking environment	4
Listener 3 (Li3)	NS	ESL teacher	No oral exam training or experience	Non-English speaking environment	4
Listener 4 (Li4)	NS	ESL teacher	Trained and experienced UCLES and SUPSI oral examiner	Non-English speaking environment	4

Figure 3. Listener characteristics

Three types of verbal report where gathered in this study. The first was *concurrent*, where the listener spoke while the recording of the interview was being played. The second was *immediately retrospective*, where at the end of a section the recording was stopped, enabling the listener to speak about what had just been heard. Finally, the listener heard the same section of the interview a second time, and the listener produced another *concurrent* report. This time however the recording was stopped whenever the listener began to speak, thus enabling the listener to elaborate. This third report was deemed necessary out of concern for the difficulty the non-teacher listeners might have in verbalising their thoughts while listening to the recordings, or in expressing themselves immediately after listening to a section of the interview. This proved in fact useful as Li2 was reluctant to speak during the non-paused concurrent parts (see Appendix C). Figure 4 outlines the data gathering procedure for Stage 2.

This procedure resulted in recordings of each Listener's verbalisations, approximately 45 minutes in length, that were then transcribed. All transcribed verbalisations for all listeners are found in Appendix C.

		Playback action	Listener	Report type
This procedure was repeated for Speakers 1,2 and 3 for all Listeners				
Part 1: Picture task	Part 1	No pause in playback	Speaks WHILE listening	Concurrent
	PAUSE	Pause playback	Speaks after listening	Retrospective
	Part 1 REPEATED	Pause playback whenever Listener begins verbalising	Speaks WHILE listening but playback paused for elaboration	Concurrent with pausing
Part 2: Topic	Part 2	No pause in playback	Speaks WHILE listening	Concurrent
	PAUSE	Pause playback	Speaks after listening	Retrospective
	Part 2 REPEATED	Pause playback whenever Listener begins verbalising	Speaks WHILE listening but playback paused for elaboration	Concurrent with pausing

Figure 4. Procedure for stage 2 data gathering

3.3. Coding of verbal protocols

The validity of any VPA is related to the accuracy and validity of the coding process (Hubbard et al 2006, in press, Green 1998: 68-71). This process is however linked to how the verbalisations are segmented and ultimately coded and categorised, since this will in the end represent our unit of analysis. Therefore the following steps were taken towards providing information as accurately as possible given the resources available for this study.

1. Since the purpose of this study was to investigate the thought processes of people as they listen to NNSs, ideally each segment or protocol should capture single thought processes (Green 1998 73-76). Identifying these provides the basis for the segments, which in turn allows for coding to take place. For example, the following excerpt from Li2

> Wow that's amazing. Well, it's amazing. I think he's a new learner of English and he was speaking in the present tense because I don't think he understands how to speak in English, the continuous, how we say “he is reading”, but he's very very observant.

was broken up into the following segments using pauses or syntactic markers as reference points.

Wow that's amazing. Well it's amazing.../
... I think he's a new learner of English.../
... **and** he was speaking in the present tense .../
... because I don't think he understands how to speak in English the continuous how we say “he is reading”.../
... **but** he's very very observant.../

2. After segmenting all of the Listener tapescripts, each was labelled. Initially this consisted of describing the process in the segment. Taking the above example, descriptions are provided in the column on the left.

listener affective reaction	Wow that's amazing. Well it's amazing.../
making inference about level of English / speaker property	... I think he's a new learner of English.../
noting the use of verb tense / grammar / form	... and he was speaking in the present tense .../
making inference about speaker knowledge / ability	... because I don't think he understands how to speak in English the continuous how we say “he is reading”.../
noting speaker property: observant	... but he's very very observant.../

3. When all segments from all tapescripts had descriptions, these were analysed, simplified and grouped into categories or themes, in order to produce a workable and organised coding system. Naturally, if categories of the code cannot be reliably applied, conclusions derived from their use will be inaccurate (Green 1998: 93-94). A first draft of the coding system was tested with another teacher, resulting in an agreement coefficient of 0.66. It was apparent that the lack of agreement was largely owing to insufficiently clear code descriptors as well as too much overlap between the codes. Codes were then reorganised and simplified, using clearer definitions. Another check resulted in higher agreement, with an interrater code-reliability of 0.89 This was deemed sufficient for the purposes of this study. The final set of main

protocol categories is given below (the complete list with sub-categories is in Appendix D).

LISTENER VERBALISATIONS OF THOUGHTS:

S REFERRING TO SOMETHING ABOUT THE SPEAKER

ST traits
SS emotional states
SE previous personal experience
SL level of English ability

P REFERRING TO SOMETHING ABOUT THE SPEAKER'S PERFORMANCE

PC content
PE manner of expression or the way content is expressed
PI interaction, communication and task completion

L REFERRING TO SOMETHING ABOUT or LOCATED WITHIN THE LISTENER

LAR Listener affective reaction
LQJ Listener qualitative judgement
LINT Listener act of interacting with speaker: by rephrasing, repeating, responding to or interpreting content or saying what speaker is doing
LTE referring to teaching experience
LC Indication of Listener comprehension or degree of Listener comprehension

T REFERRING TO SOMETHING ABOUT THE TASK

listener referring to some property of the test/interview task

I REFERRING TO SOMETHING ABOUT THE INTERVIEWER

listener noting some property of the interviewer

O REFERRING TO OTHER REFLECTIONS OR SPECULATIONS

NS qualities in general, NNS qualities in general and other reflections

Figure 5. Main categories of verbal protocols

The S category refers to think-aloud protocols where the listener made reference to something about the speaker. Examples include comments about a speaker trait or emotional state. The P category, is in reference to protocols that mention some performance feature. This refers to when listeners noted the content of what the speaker was saying or the manner in which content was being expressed. The L category indicates when listeners referred to something going inside themselves or mentioned something about themselves. These included some explicit emotional reaction (LAR), a judgement of quality (LQJ) or some reference to the degree of comprehension of the speech samples (LC). For example, LC was the coding for statements such as “I don't understand”, “I think I understood what he meant” or “I had no trouble understanding.” LINT requires some explanation. On occasion, a listener would comment in a manner that appeared to mimic the act of interacting with the speaker, by interpreting or commenting on comment or even responding to the unseen speaker, despite the fact that interaction was impossible since the listener was listening to a recording. This protocol type was defined as listener interaction with the content of the speech samples. Merely noting content in speech (PC) and LINT, where in some sense the listener contributes to the speaker's message, appear to represent two types of reaction to content. To illustrate the difference let us look a 2 examples.

Li4-051) PC ... and talks about the consequences of certain actions .../

Li1-062) LINT That's true, yeah that's very true, and you really.../

In the first example, Li4 limits himself to simply stating what the speaker is talking about, without actually reacting to that content. In contrast Li1 is not detached from what is said, and, as though engaged in conversation, interacts with the content. Acts of interpretation and paraphrasing were also considered examples of LINT where listeners attempted to “get into” the content as opposed to simply noting its presence. Some but not all main groups have sub-codes. These are important when a more

detailed analysis is required, as we will see in the next chapter. For example the code group P, that is when listeners noted some aspect of the speaker's performance, actually consists of many sub-codes. These sub-codes were put under 3 main headings. These referred to 1) when listeners reflected on what the speaker was trying to communicate or content (PC), 2) the way or manner in which the speaker tried to communicate or expressed him/herself and the tools needed to do so (PE) and finally 3) when listeners noted some aspect of interaction in the performance (PI). If, for instance, a listener said “some grammatical errors”, this was coded as PE, since it refers to how something was expressed or the tools needed to express some content. If on the other hand the comment was “she's talking about her free-time activities”, this would be coded PC since the raw content of what is being communicated is attended to. A comment like “he's trying to respond to the questions” would be coded PI, since this refers to interaction rather than content or how content is expressed. A full list of all codes and sub-codes is provided in Appendix D.

4. Protocols were then labelled again, using the above system and numbered as in this example.

001)	LAR	Wow that's amazing. Well it's amazing.../
002)	SL	... I think he's a new learner of English.../
003)	PE	... and he was speaking in the present tense .../
004)	SL	... because I don't think he understands how to speak in English the continuous how we say “he is reading”.../
005)	ST	... but he's very very observant.../

The full coding of all transcripts for all listeners is found in Appendix C.

Chapter 4

4.1. Data analysis

It must be remembered that the following is entirely based on what listeners were able to verbalise when listening to the speech samples. Naturally, it is possible that listeners noted other features or reacted to speech implicitly or unconsciously without verbalising. Other tools would be necessary to explore these areas. Nevertheless, the VPA should give us an idea of how listeners think, and in a way that under normal circumstances would not be verbalised at all. Protocol listener data is presented in graphs below and in Appendix E. The listener characteristics found in Figure 3, with Li1 and Li4 representing the two "poles" of our study, should be kept in mind, as this will assist the reader through the following analysis.

4.1.1. Quantitative differences in individual listener cognitive activity

In analysing the number of listener protocols that refer to different types of mental activity, we notice that each listener exhibits a distinct pattern. For example, for language samples from all 3 speakers, Listener 1 consistently verbalised a relatively high percentage of thoughts regarding speakers' properties (S codes), a much lower amount pertaining to speakers' performances (P codes), and a relatively high percentage of verbalisations regarding the effect the speech samples had on the listener herself, in particular comprehension (L codes). The pattern is manifest regardless of the speech samples. This can be clearly seen in Figure 6 where the percentages of Li1's protocols for each protocol type are compared for S1, S2 and S3. This is significantly different from Li4, whose profile is the exact opposite (Figure 9), where most of Li4's protocols focus on speakers' performances (P codes) and very little on inferences regarding the speakers (S codes) or on the effect speech had on the listener himself (L codes). Li2 and Li3 both exhibited different but consistent patterns (see figures 7 and 8), suggesting that different listener types are emerging, each tending to focus on particular areas of performance. Looking at sub-codes in Appendix E, we can see that Li1 had a very high number of Speaker Trait (ST) and

Listener Interaction (LI) protocols. Li2 focused most on grammar (PE gr) and vocabulary (PE voc) while Li3 most frequently verbalised regarding the effect the speech samples had on her (LAR and LC) and pronunciation (PE pro). Li4 focused mostly on the content of the performances and the manner of expression (PE). Clearly, each listener had a specific area of interest that led to more verbalisations. The fact that speech triggered thought processes that were generally the same over all speech samples, despite differences in the speech samples, suggests that listener perceptions were shaped by their own cognitive patterns rather than the performances, otherwise different performances would lead to different protocol patterns within individual listeners. This also implies that the same language production will not necessarily cause the same areas of thought to be triggered in different listeners, as is demonstrated in Figure 10 where the listeners' reactions to S1 are compared.

Figure 6. Li1 reactions to Ss (%)

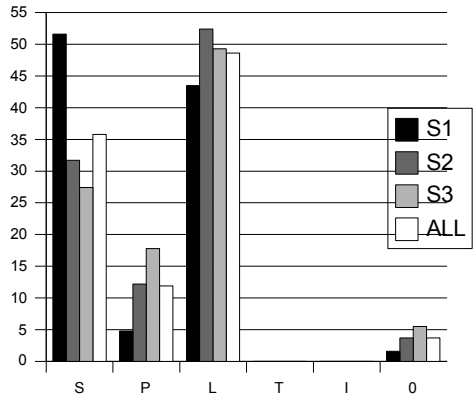


Figure 7. Li2 reactions to Ss (%)

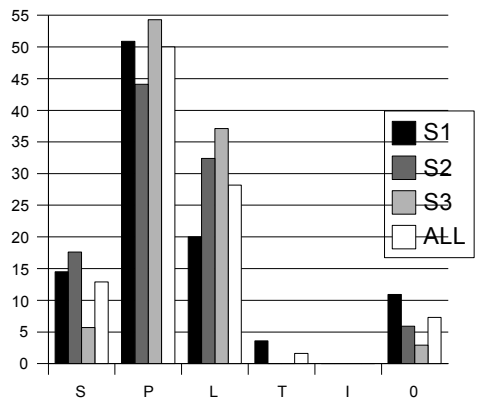


Figure 8. Li3 reactions to Ss (%)

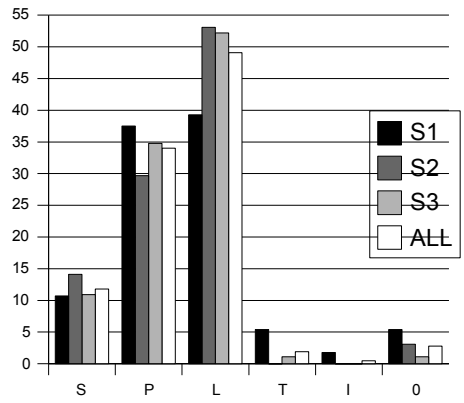


Figure 9. Li4 reactions to Ss(%)

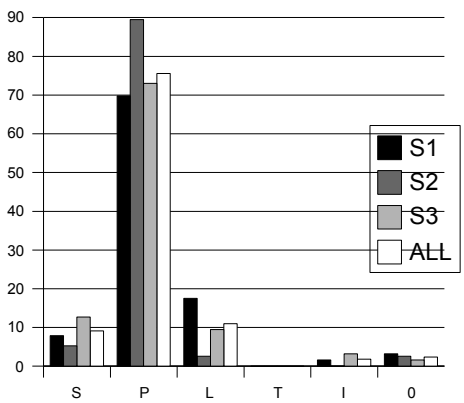


Figure 11 compares the protocols of non-teachers and teachers in the main protocol categories for all speakers. The teacher focus on language performance (P) in contrast to the non-teacher focus on speaker-centred thoughts (S) or listener-centred thoughts (L) can be seen from the chart. This implies that a teacher's thoughts are dominated by performance, correct language use and analysing the way speech is used, while a non-teacher's thoughts are more concerned with what is being said, drawing inferences about the speaker, the effect language has on her/himself and comprehension. As discussed above, these patterns are more marked in the two extreme positions in our study, in the protocols of Li1 and Li4.

Figure 10. Listener reactions to S1 (%)

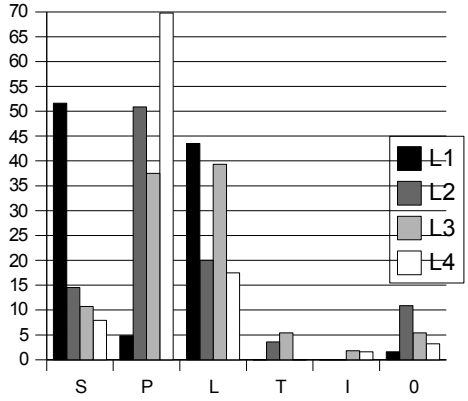
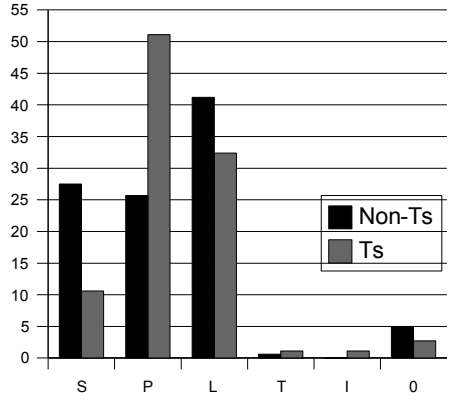


Figure 11. Listener reactions to all Speakers



It is suggested that these cognitive patterns are shaped by personal experience and environment. Li4's experience as a trained oral examiner appears to have led to an almost exclusive focus on performance features and the suppression of other thoughts regarding the speaker and personal reactions to the content of speech (Figure 9). This of course matches the experience of a trained assessor, who must focus exclusively on assessment criteria rather than other areas. Although asked to not "evaluate" the speakers but rather to simply verbalise what came to mind naturally, Li4 appeared to take on the role of evaluator throughout. 124 codings out of 164 in total, or 75.6%, were in the Performance category (how content is expressed).

Experience also appears to have affected Li2's protocols. Li2 produced the highest percentage of PE gr and PE voc codings of all raters (18.5% and 12.9% respectively).

These two categories were also the most frequent of all of Li2's codings. This attending to grammatical and vocabulary accuracy is likely to be caused by Li2's experience of living in a non-English speaking environment and having learned several foreign languages. More frequent exposure to the speech of NNSs than Li1, and FL learning experience, may account for the increased saliency of these areas. In contrast Li1 has no experience of learning a foreign language, and lives in an English speaking environment. As a result, there is less focus on P codes in general in Li1's protocols. P codes account for only 11.9% of Li1's protocols compared with 50% for Li2, 34% for Li3 and 75.6% for Li4 respectively (Appendix E). In contrast Li1's focus on inferences about the speakers and comprehension appears to match that of someone without FL classroom experience as a learner or much exposure to learners of English. In encountering a NNS, such a person would not analyse the NNS's speech. Perhaps some obvious features of speech that distinguish the NNS from a NS would be noticed, but the focus would be on comprehension, interpretation and interaction with the speech content as opposed to focussing on detailed aspects of language performance. This is precisely what is seen in Li1's protocols, as evidenced in the high number of S, LC and LINT codings (for a discussion of the LINT code, see section 3.3).

The ratio of PC and LINT protocols (Figure 12) further demonstrates the difference between the trained-teacher and the other listeners in our study. Since PC refers to when listeners note content with no personal "interaction" with that content, and LINT to when content is in some way interacted with, this ratio is a good indicator of what listeners do with input. The marked high and low LINT to PC ratios in Li1 and Li4 respectively appear to support our previous discussion regarding how assessment training results in the suppression of the natural inclination to interact with what a person is saying.

	Li1	Li2	Li3	Li4
LINT	53	8	46	1
∑ PC	1	7	13	37
LINT:∑PC	**53.0**	**1.14**	**3.54**	**0.03**

Figure 12. Comparison of LINT codes and PC codes

The implications of the above points are important. If true that experience strongly influences cognitive activity and thereby perceptions of language performance, as our findings seem to support, the complete dominance of current models (such as the CEFR) by the teacher perspective represents a serious weakness to their usefulness to non-teachers. This is further seen in the degree of fit between the naturally produced reactions to language samples found here, and the assessment criteria used in operationalised GP testing. Figure 13 shows UCLES speaking test assessment criteria (UCLES 2001: 48), the code category under which the criterion would be located in this study, and whether the UCLES criterion was found in the protocols of each listener, Y for yes, N for no. If the criterion was found in a listener's protocols, an example protocol line is given in brackets for reference.

As can be seen in Figure 13, there is a pattern from Li1 to Li4, where the number of matches with the UCLES criteria steadily increases, in support of the experience hypothesis. Owing to Li1's lack of FL study experience, a low correspondence between Li1's protocols and UCLES criteria was found. Li2's protocols show more correspondence as discussed earlier, likely due to Li2's FL learning experience and environment. Li3, experienced as a teacher, but lacking assessment training, showed a greater (56%) match with the UCLES criteria. In contrast, Li4's protocols matched nearly all of the possible UCLES criteria. Li4 made use of a broad range UCLES criteria, which is especially interesting given the relatively short speech samples. The point here is that the natural occurrence of UCLES criteria appears to diminish moving towards the listener that does not have a comparable environment or experience (towards Li1). This also suggests the teacher perspective possibly entails analytical thought processes that are unnatural to non-teachers.

UCLES speaking assessment criteria	Code where this competence is located	Li1	Li2	Li3	Li4
Grammar and vocabulary					
grammar accuracy	PE gr	N	Y (018)	Y (003)	Y (06)
vocabulary appropriacy	PE voc	Y (095)	Y (0107)	Y (0133)	Y (066)
vocabulary range	PE voc	N	Y (087)	N	Y (0111)
Discourse management					
grammatical range	PE gr	N	N	N	Y (057)
coherence	PC cont or PE flow	Y (0184)	N	Y (013)	Y (071)
extension	PE ext	N	N	Y (015)	Y (0110)
Pronunciation					
sounds	PE pro	Y (025)	Y (014)	Y (033)	Y (03)
linking of sounds	PE pro	N	N	N	Y (01)
intonation	PE sty	N	N	Y (078)	Y (037)
stress	PE pro	N	N	N	N
Interactive communication					
speed	PE flow	Y (069)	Y (010)	N	Y (065)
rhythm	PE flow	Y 0174)	Y (0110)	Y (035)	Y (0149)
maintaining: turn taking	PI	N	N	N	N
maintaining: initiating	PI	N	N	N	N
maintaining: responding appropriately	PI	N	Y (067)	Y (074)	Y (071)
repair strategies	PE flow sc	N	N	Y (0165)	Y (081)
task development	PE ext	N	N	Y (0176)	Y (0118)
willingness to develop	ST	N	N	N	Y (011)
Total Y		**5**	**7**	**10**	**15**
Match with UCLES categories		**0.28**	**0.39**	**0.56**	**0.83**

Figure 13. UCLES speaking assessment criteria found in listener protocols

Although we cannot generalise broadly to entire population groups because of the small number of participants, our findings suggest the existence of different listener types, with distinct cognitive patterns based on their personal experience. This raises important questions. If assessment criteria cannot empirically be shown to exist naturally within a TLU, or better, within a context of test score interpretation, can they be considered authentic? If not, should they be used? How meaningful are

such assessment criteria to those whose thought processes do not allow them to occur naturally? Is it valid to impose assessment criteria, which are based on teacher experience, training, and attendant patterns of cognitive processes, if these do not match those who will interact with test-takers in post-test situations? The weak match between the overall cognitive activity of teachers and non-teachers, especially between the non-teacher NS with no FL experience living in an English-speaking environment, and the trained-teacher in the use of UCLES rating scales, as found here, suggests that the answers to the above questions are decidedly negative. This must be cautiously stated however, as it would be necessary to repeat the same procedures over a broader range of tasks and testing conditions with more participants (testees and raters). The TLU simulated in this study is by no means representative of all general proficiency target language contexts. If the same weak match is revealed in further studies then we may be more confident in our conclusions.

4.1.2. Qualitative Differences in Cognitive Processes

The above sections compared the quantities of protocols, for the purpose of finding patterns in the way listeners react to speech. A considerable difference was found in the types of mental activity stimulated in our listeners given identical speech samples. Now we will explore qualitative data and discuss some interesting findings. While there are many items of interest, space will not permit an examination of every one. Guided by our research questions, we must instead focus on some of the more interesting findings directly related to our objectives. We will begin with a comparison of protocols from Li1 and Li4 in reaction to S2 (CEFRL level A2) and S3 (CEFRL level C1).

S2					
Li1			Li4		
068)	PE flow	And he's uh speaking slowly./	075)	PE flow	He's speaking very slowly .../
072)	SL	Yes, he struggles a little bit with his English.../	076)	PE flow	... um searching for almost every word .../
070)	PE flow	Thinking for his uh, the words to use./			
091)	SL	He struggles with his words.../			
069)	LC	Easy to understand./	080)	PE flow	Um because of the way he's speaking every word is completely separate there's no fluency at all it's almost like a machine producing a list of words .../
073)	LC	... but it's really clear, very clear./			
092)	PI	... but he uh he really is able to express himself./			
077)	LQJ / PE sty	He describes everything really well./			
078)	LC	Because I can follow him with the pictures./	095)	O	I I think he would have a lot of difficulty communicating with people in normal situations because most people would not have the patience to listen long enough to hear even one sentence./

Figure 14. Selected Li1 and Li4 protocols compared (A)

We can see above that Li1 and Li4 both note the same aspect of S2's performance: slow speech, and S2's difficulty in finding the words to say. The similarity ends however as the two listeners process these aspects in different ways. Li1 focuses on the fact that S2's slow speech is clear and easy to understand. Li1 says that S2's speech flowed better than S3's speech (Li1-0174) in stark contrast to Li4's opinion that fluency was absent in S2 (Li4-080) but S3 is very fluent (Li4-0159). It appears that Li1's focus is on following and understanding S2's speech, and given that content of speech is successfully transmitted, this leads to a positive view. In other words, Li1's view is shaped by the priority of comprehension. The limited language resources are noted, but are of less importance and contribute less to the overall impression. This is further supported quantitatively as Li1's codings show

(Appendix E and discussed above). In contrast, Li4 appears to have a different priority. In fact Li4 makes no mention of being able to understand S2, instead, focussing almost exclusively on linguistic knowledge as expressed in performance, describes the slow speech as “plodding” (097), “machine-like” (080) and without emotion (098), as opposed to “clear” and “easy to follow” which is Li1's view (see above). Li4 makes an interesting observation that non-teachers would have considerable difficulty communicating with S2 because of S2's limited language resources, and slow plodding style (076) which would be difficult to tolerate. We can only assume that if Li1 found S2 easy to follow and understand, and his speech clear, descriptive and expressive, as clearly evidenced in Li1's protocols, it would not be as difficult to establish and maintain communication as hypothesised by Li4. As we shall see momentarily, Li1 in fact expresses intolerance not for S2's slowness, but for S3's fast speech which is also the opposite of Li4's view. Li4's speculation regarding non-teachers' tolerance is contrary to the implications of Li1's comments. It is suggested that personal cognitive patterns, owing largely to experience and environment are the cause of this difference in how not only the same language samples are processed, but even the same speech characteristic (in this case slow speech).

Li4 has been trained to focus on a specific set of criteria and reach conclusions based on careful consideration of those criteria. During the first phase of S3's speech samples (Li4-0100-0116), Li4 clearly goes through a check-list of performance features that result in a conclusion. Here Li4 notes good fluency, good task fulfilment, extensive vocabulary, some minor errors, but a good range of verb tenses, good extension, good intonation, expressiveness and finally the correct usage of the 2nd conditional, all leading up to the conclusion:

0116) SL So he seems to have a good level of ability./

In other words, the presence of a specific broad range of discrete elements of performance, which if compared directly with UCLES criteria (section 4.1.1.) clearly

demonstrate the effect of Li4's training and experience, all contribute to the final inference regarding proficiency. A rather different effect occurs within Li1 however.

S3					
Li1			Li4		
0155)	LC	I can't, I'm not understanding him really well here./	0108)	LC	... but nothing that's getting in the way of communication./
0164)	LC PE flow	I I had difficulty understanding, uh comprehending what he was uh saying because he was speaking quicker and uh, I have to listen to it again now./	0134)	PE flow	... he expresses himself in a very natural way uh pausing, emphasising./
0156)	PE flow	He's rushing with his thoughts./			
0172)	LC	And um, OK. I wonder, I'm, I'm, I still am not sure really where he's going with this./			
0167)	LC	...but um, somehow I have difficulty understanding where he's going./	0132)	PE flow	He's speaking very very easily and fluently./
0168)	PE flow	... because it doesn't, to me it's it's not flowing well./	0159)	PE flow	... and you know, speaks very fluently./
0183)	PE flow	... and I'm, so, and it seems to be, the conversation, the interview is very scattered./			
0173)	PE flow	The other 2 were uh, they had a rhythm to their uh, to the interview./	0149)	PE flow	He also has a kind of natural rhythm that I think of as being uh, very British English./
0174)	LC	I'm have, I'm just having a little difficulty./			

Figure 15. Selected Li1 and Li4 protocols compared (B)

In Figure 15, it can be seen that Li1 had considerable difficulty understanding S3, in fact the most difficulty of all speakers, whereas Li4 clearly had no difficulty at all. Li1 identifies S3's faster speech as the cause of this difficulty, and yet this faster speech was actually taken as a sign of fluency in Li4:

0150) O ... where people will speak very slowly and then they'll put a whole bunch o' words together real fast. This to me is a typical educated British way of speaking .../

This faster speech understood as “natural” and “speaking easily” by Li4. Interestingly, Li1 is not alone in processing S3's faster speech differently from the trained-teacher in our study. The other non-teacher, Li2, interprets S3's faster speech as a sign of nervousness (Li2-0105, 0106). Li1 (0168, 0173), in direct contrast to Li4 however (0134, 0150 and 0159), says that S3's speech is not fluent, and less so than the other speakers (Li1-0173). The mixture of speed, unclear speech and perhaps pronunciation difficulties (pronunciation is not mentioned explicitly by Li1, but is noted extensively by Li2 and Li3), apparently resulted in Li1's inability to understand or follow S3's speech. Many of the discrete and clearly teacher-experience related aspects of performance go unnoticed by Li1, and do not contribute to any opinion of S3's performance. It seems that once again, Li1's overall priority of trying to understand S3 is rooted in the layperson's experience and a listening pattern that appears to be holistic rather than based on discrete elements “summed up”. Significantly, Li2 (083 and 093) also noted difficulties in understanding S3, but these did not lead to an overall negative impression. Li2's experience with NNSs and attendant expectations and increased tolerance for difficulties in comprehension may have played a key role in Li2's rating of S3's performance as the best of the three. Li1 was unable to resolve the difficulties and this resulted in Li1's rating S3 as the worst of the three (see Figures 16 and 17 below).

In addition to the significant differences in the categories of mental activity of our listeners, qualitatively and quantitatively, it seems apparent that there is a difference in the value or weighting some categories have in different listeners. This prioritising of certain aspects has strong implications regarding the weighting of assessment criteria leading to tests scores. GP testing traditionally gives equal value to each criterion in rating scales, as, for example, UCLES tests assign equal value to grammar and vocabulary, discourse management, pronunciation and interactive communication (COE 2000: 195). Empirical evidence is not offered in support of this stance. Instead, only an appeal to reason is given, that “a candidate’s aggregate score over the whole range of language skills is the most appropriate measure of ability for exams of this type and purpose.” (Saville 2003: 107). The equal value approach was evident in Li4's protocols, but was not supported in the case of Li1.

The non-teacher without FL learning experience in our study placed greater value in Interactive Communication and perhaps Pronunciation, to use the UCLES categories.

Speaker	Li1	Li2	Li3	Li4	Avg non-teachers	Avg teachers
S1	1	1	2	2	1	2
S2	3	3	3	3	3	3
S3	1	1	1	1	1	1

Figure 16. Ranking of speaker language ability

Speaker	Li1	Li2	Li3	Li4	Avg non-teachers	Avg teachers
S1	2	3	2	2	2.5	2
S2	1	2	3	3	1.5	3
S3	3	1	1	1	2	1

Figure 17. Ranking of speaker performance

It also appears that the priority certain features of communication have in the minds of listeners because of experience is key to understanding how listeners arrive at opinions regarding ability and performance. In Figures 16 and 17 above, the rankings of speaker ability and performances are given for each listener. There are several interesting differences. First of all, neither non-teacher listener was able to distinguish between S1 and S3 in terms of ability. Both teachers were able to make a distinction ranking S3 ahead of S1. This is expected given the training of Li4 and the Li3's experience of teaching, organising classes according to level, selection of course materials and so on. Second, there appears to be a tension between the perception of language ability and the parallel quality of performance that is supposed to be the act of putting the ability into practice. The distinction between the two is strong in the non-teachers, but lacking completely or perhaps suppressed in the teachers. Both teachers agreed on the ranking of performances, and significantly, that the performances matched their opinions of ability. This could suggest that, for teachers, the two are linked. It is more likely however that greater sensitivity to a broader range of discreet points in both teachers (see Figure 13

above) results in a common judgement. In contrast, neither non-teacher ranking of performances matched their perception of language ability. It appears that language ability and performance were more easily differentiated suggesting that non-teacher ideas relating to performance are different from teachers. Li1 and Li2 agreed that S2 had the lowest level, but on average gave S2 the highest ranking in performance. Clarity of speech, effort and managing to communicate despite a low level of language knowledge seem to have won out in their view. Naturally, the quantitative value of these rankings is extremely limited because of the small number of participants. Nevertheless, the differences are striking when put in the context of the qualitative analysis presented earlier.

Summing up, our findings appear to show a considerable qualitative difference in listener cognition, again apparently based on experience, leading to different priorities in forming perceptions about NNSs. Thus, all research questions have been answered negatively: no strong quantitative or qualitative match was found between teacher and non-teacher listener cognition, leading to the final conclusion that the construct of GP as operationalised in standard GP testing frameworks does not adequately represent the non-teacher perspective.

Chapter 5: Conclusions and recommendations

Recently Brown, Iwashita and McNamara (2005) performed a rater cognition study of university-based ESL and communications skills experts, based on the assumption that their study would allow for the distillation of the expert perspective within a specific context, thus permitting the creation of an authentic and unified rating scale appropriate to the context of use. This idea of identifying the expert perspective obviously hinges on the assumption of homogeneity among these experts, and can only be found given their common experience as experts. It appears that owing to a lack of homogeneity of experience among the participants in our study, considerable differences were found between NS teacher cognition and that of non-teachers. Marked differences were found in particular between our teacher trained in the use of rating scales and our non-teacher with no FL learning experience and living in an English speaking environment. While conclusions must be cautious, our findings suggest that:

- teachers and non-teachers attend to different features of language performance based on their experience: non-teachers are more concerned with comprehension, interpretation and interaction, whereas teachers are more interested in assessing a broad range of performance features

- teachers, owing to classroom and possible rater training experience, show sensitivity to features of speech and performance that non-teachers do not attend to, suggesting that some classroom schemata-based criteria may not be meaningful or useful to non-teachers

- different listening priorities in teachers and non-teachers lead to different levels of tolerance, affecting perceptions of performance. Non-teachers show greater tolerance of linguistic inaccuracies in favour of comprehension while teachers show greater tolerance of comprehension difficulties in favour of a broad range of assessment markers

- teachers and non-teachers at times interpret the same feature of speech, for example rate of speech, in quite opposite ways.

These differences suggest that current GPT models do not adequately represent the perceptions of non-teachers. It has been argued that GPT should be informed by

consulting those that ultimately represent the context in which language will be used and scores interpreted, rather than teachers or language experts. Since assessment criteria based on the cognitive patterns of those that interpret test scores will have greater authenticity and result in more meaningful scores, it is suggested that GPT assessment frameworks be redesigned in recognition of the differences in teacher and non-teacher cognition. Tentatively, our findings point to the need for assessment criteria that assign greater value to comprehension- and interaction-enabling features of speaking ability, such as clarity of speech, speed or pronunciation as perceived by non-teachers, as opposed to a broad range of theoretical, teacher-schemata based criteria that do not occur naturally in non-teachers' thoughts.

It is suggested that this small-scale exploratory study should be followed by larger scale studies, applying the principles and methodology outlined here and using statistical tools for more accurate population-representative conclusions. These studies should examine listener and co-communicator cognition over:

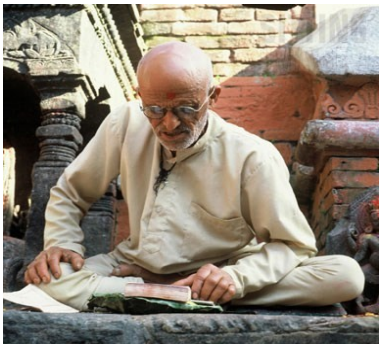
- different interactional contexts, such as group interaction, monologues, presentations and so on
- a full range of task types
- a full range of task difficulty levels
- greater numbers of participants from a broader range of experiential groups to allow for conclusions to be made across populations.

Findings from these studies could then lead to the definition of assessment criteria that would potentially answer many of the problems the general proficiency testing industry is now faced with. By developing assessment criteria based on the perceptions of those that are found in the context of test use, we address the problem of authenticity and construct validity more effectively. This would in turn increase the possibility of providing meaningful information to stakeholders, better addressing the ethical issue of stakeholder priority. The *a priori*-based frameworks now in use, that are rooted in an expertise that is limited to the experience of the classroom, could then be replaced by a truly empirical framework, useful and meaningful to the general public.

Appendix A. Elicitation instruments

1. Picture Task

Look at these pictures and talk about them in any way you like.



2. Discussion

"I have a job so I can put food on the table, but my passion is what I do in my free time." Is this the right way to think about work and life? Let's discuss this together.

Appendix B. Speech sample scripts

Legend:
I: Interviewer
S: Speaker

Speaker 1: Picture Task

S1:	YEAH THIS PICTURE THIS PICTURE I SEE DIFFERENT GENERATION / FOR EXAMPLE I SEE VERY YOUNG TRYING THAT TRYING TO LEARN COMPUTERS / AND THEN I SEE OLD PEOPLE / IF YOU CAN SAY THAT / TRYING TO KEEP READING KEEP TO TO STAY IN OUR SOC(IETY) / NOT TO BE (BE)HIND AND SOCIALISE PEOPLE LIKE YOUNG PEOPLE / THEY STAY TOGETHER TALKING ENJOY THEMSELF / AND THAT'S WHAT I'M THEM I SEE / YEAH PEOPLE THAT TRY TO TO BE BUSY TO WITH THEIR LIFE AN' EVERYTHING TRY TO KEEP GOING, TRY NOT TO STAY BEHIND THE TO THE / HOW CAN I SAY / TO THE IN THE FUTURE NOW JUS(T) TO KEEP GOING NOT TO LOSE / UH UM YES JUST TO IN THE PRESENT YEAH
I:	CAN YOU SEE ANY SIMILARITIES OR DIFFERENCES IN THE PICTURES?
S1:	WELL I SEE TWO PE(OPLE) TWO OLD PEOPLE THAT TRY TO TO READ IN SOME SOME BOOKS / THEN I SEE YOUNG PEOPLE THAT TRY TO LEARN COMPUTERS / AND THEN ANOTHER GROUP OF PEOPLE THEY TRY TO SOCIALISE TO TALKING ABOUT UM IN GENERAL STUFF / SO ANYTHING ABOUT YEAH JUST STAY TOGETHER AND ENJOY LIFE
I:	UH HUH UH HUH ANYTHING ELSE THAT UH COMES TO MIND?
S1:	MM NO AT THE MOMENT NO I DON'T I DON'T SEE ANYTHING THAT COMES TO MIND NO
I:	UM DO YOU UHHH USE A COMPUTER OFTEN?
S1:	YES I DO OFTEN YES UH HUH
I:	AND DO YOU THINK THIS IS A GOOD THING OR…?
S1:	WELL IT IT DEPENDS WHAT YOU WHAT YOU WHAT YOU MEAN
I:	WHAT'S YOUR RELATIONSHIP LIKE WITH COMPUTERS?
S1:	WITH COMPUTER I LIKE UH BUT NOT VERY MUCH BECAUSE UH I LIKE TO USE INTERNET WHEN I NEED TO WRITE SOMETHING IS VERY USEFUL / BUT I THINK COMPUTER TRY TO UM PUT PEOPLE TO ISOLATE EACH OTHER / I DON'T LIKE VERY MUCH BECAUSE I PREFER TALK WITH PEOPLE BECAUSE TO TO MAKE RELATIONSHIP / BECAUSE COMPUTER IS NICE CAN DO A LOT OF STUFF BUT THE COMPUTER TRY TO LET YOU TO BE YOURSELF WITH YOURSELF AND NOT TO COMMUNICATE WITH PEOPLE FACE TO FACE
I:	UH HUH, UH HUH
S1:	SO THAT IS I THINK IS IN THE FUTURE CAN BE COMPLICATE BECAUSE PEOPLE TRY TO TO NOT TALK TO EACH OTHER BUT JUST IN DISTANCES SO THAT'S CAN BE A PROBLEM MAYBE I DUNNO / THAT'S MY POINT OF VIEW BUT I PREFER TO TALK WITH PEOPLE DON'T USE COMPUTER ALL THE TIME

Speaker 1: Discussion

I:	IS THIS THE RIGHT WAY TO THINK ABOUT WORK AND LIFE? LET'S DISCUSS THIS TOGETHER.
S1:	TO LIVE WITH / WELL UNFORTUNATELY WE HAVE TO WORK ANYWAY / SO IF YOU DON'T FIND A JOB THAT YOU REALLY LIKE UNFORTUNATELY WE HAVE TO DO BECAUSE IF WE DON'T HAVE A WORK WE CANNOT SURVIVE / WE CAN DO

	ANYTHING AND YOU CANNOT ENJOY AS WELL YOUR FREETIME / SO IF IN THE MEANTIME YOU CAN DO THIS JOB THAT YOU DON'T REALLY LIKE IT BUT YOU CAN LOOK IT FOR JOB THAT YOU VOULD LIKE TO DO / AND SO AND SAME IN THE MEANTIME YOU CAN ENJOY YOUR FREE TIME AND THEN ...
I:	SO DO DO YOU THINK THAT IT'S NECESSARY TO FIND A JOB TO BE HAPPY NO / TO BE TO BE A HAPPY PERSON IS IT NECESSARY TO FIND A JOB WHICH IS ALSO YOUR PASSION / OR IS IT ENOUGH TO SAY, EH, UHHH, MY PASSION / IT'S OK IF I DO THIS IN MY FREETIME AND I ACCEPT A JOB WHICH MAYBE IS NOT SO INTERESTING BUT / YOU KNOW / I NEED TO EAT RIGHT SO UH UH MMM …
S1:	WELL TO HAVE A JOB I THINK YOU HAVE TO LIKE IT BECAUSE AFTER A WHILE YOU BECOME TO HATE THIS JOB.
I:	MM HMM
S1.	AND I THINK IF YOU DON'T LIKE THE JOB IT WILL AFFECT AS WELL YOUR FREE TIME / SO IS MOST IMPORTANT IF YOU CAN FIND SOMETHING THAT YOU LIKE IT YOU CAN ENJOY SOMETHING WH(AT) WHAT YOU DO IN YOU HAVE A LITTLE SATISFACTION WHAT YOU YOU DO IN DURING YOUR WORK TIME AND THEN YOU CAN ENJOY MUCH BETTER YOUR FREE TIME
I:	UM DO YOU HAVE ANY PASSIONS / SOMETHING THAT YOU WOULD REALLY REALLY LIKE TO DO FOR A JOB?
S1.	WELL MY PASSION IS TRAVEL / SO YES UH IF I CAN (LAUGHS) I CAN TRAVEL ALL MY ENTIRE LIFE THAT'S THAT'S MY PASSION YES / I'M I'M BEEN MM NOW WORKING IN THE TOURIS, SO THAT'S I'M HAPPY / BUT UNTIL TODAY I HAVEN'T FIND THE REALLY JOB THAT I'M SUITABLE THAT I REALLY LOVE IT SO MUCH / SO I'M TRYING EVERYTHING NOW I'M TRY TO WORK FOR THIS HOTEL / FOR NOW I LIKE IT THEN I DUNNO / IF I KEEP GOING IN THIS FIELD OR MAYBE I WILL CHANGE OR I DUNNO REALLY NOW I'M TRY TO TAKE THIS ONE
I:	ARE THERE CHANCES TO / I MEAN DO YOU HAVE THE CHANCE TO TO TRAVEL WITH THIS KIND OF WORK THAT YOU'RE DOING NOW?
S1:	NO ACTUALLY WELL YOU CAN WORK F(OR) LIKE FOR A SEASON THEN I'M I'M WORKING UNTIL THE END OF (OC)TOBER THEN I CAN GO WORK SOME OTHER HOTELS UM WHERE / I WANT TO GO YES IF I FIND A PLACE.
I:	IS THIS YOUR FIRST JOB WORKING IN A HOTEL?
S1:	YES THIS IS MY FIRST JOB IN HOTEL YES
I:	UM SO D(O) YOU THINK THAT THIS WILL HELP YOU TO FIND UH A JOB MAYBE IN ANOTHER COUNTRY UH…?
S1:	MAYBE I CAN FIND AN IN ANOTHER COUNTRY I CAN WORK IN EUROPE SO UH I I CAN FIND SOME OTHER JOBS I GUE(SS)
I:	MM HMM AHH DO YOU HAVE ANY OTHER THINGS THAT YOU LIKE DOING IN YOUR FREE?
S1:	DURING MY FREE TIME I LIKE TO GO MMM SEE MOVIES, READING, RIDE SOMETIME BICYCLE, UH THAT'S UH YES / I'M SPEND MY TIME MYSELF UH WALKING OR DO SOME JOGGING / YEAH THAT'S / VERY SIMPLE THING JUST UH
I:	ALRIGHT, THANK YOU VERY MUCH FRANCA.

Speaker 2: Picture Task

S2:	I LOOK A ONE MAN AND UH HE READ A BOOK.
	I LOOK A OTHER MAN IN OTHER PICTURE AND UH HE SEARCH THE INFORMATION FROM ONE BOOK AND THEN HE WRITING IN YOUR PERSONAL COMPUTER THIS INFORMATION.
	I LOOK A GROUP OF YOUNGER / HE SH UH THEY ARE OUTSIDE AND THEY SPEAK AND UH PLAY A COMPUTER GAME AND UH
	I LOOK A AND OTHER PICTURE UH WHERE THE YOUNGER ARE INSIDE AND UH THEY PLAY A COMPUTER GAME OR UH LEARN INFORMATICA / I DON'T KNOW
	THE DIFFERENT BETWEEN THE YOUNG OLDER MAN IS THAN ONE MAN IS OUTSIDE YOUR HOME AND UH OTHER OLDER MAN IS INSIDE IN UH A LIBRARY.
I:	IS THERE ANYTHING ELSE YOU WANT TO TO SAY?
S2:	YES / (STOPS)
I:	THANK YOU

Speaker 2: Discussion

I:	WH… D… YOU PLAY FOOTBALL RIGHT?
S2:	YES YES
I:	ARE YOU REALLY GOOD AT FOOTBALL?
S2:	NO
I:	SO THERE'S NO EH THERE'S NO POSSIBILITY THAT YOU COULD BECOME A PROFESSIONAL FOOTBALL PLAYER? YOU DON'T THINK SO? (OVERLAPPING)
S2:	NO BECAUSE I DON'T HAVE A TALENT / I I PLAY FOOTBALL WITH MY FRIENDS
I:	UH HUH
S2:	JUST FOR UH FOR HOBBY FOR RELAX
I:	MM HMM, MM HMM.
S2:	BECAUSE THE SPORT IT'S IMPORTANT IN IN LIFE.
I:	SO YOU CAN DO THIS AS A HOBBY…
S2:	(LAUGHS)
I:	… AND YOU'RE HAPPY WITH WITH THAT AS A HOBBY.
S2:	I I I HAVE UH ONE HOBBY FOR EXAMPLE CHESS UHH AND UH I THINK THAT I WANT UH UH A PROFESSIONAL…
I:	AH
S2:	… JOB
I:	SERIOUSLY?
S2:	YES BECAUSE I I I LIKE CHESS AND I I STUDY CHESS WHEN …
I:	SO THIS IS A PASSION FOR YOU (OVERLAPPING)
S2:	YES YES WHEN I HAVE A FREE TIME AND UH I I OPEN THE CHE.. CHESS BOOK AND UH LOOK UH THE POSITIONS STUDY AND UH I I PLAY A LOT THE CHESS

I:	UH IS IT POSSIBLE TO DO THIS AS A PROFESSION THOUGH / UH UH IS IT POSSIBLE TO BECOME A UH PROFESSIONAL …
S2:	MAYBE
I:	… TOURNAMENT PLAYER?
S2:	MAYBE MAYBE
I:	AND THIS WOULD BE YOUR PASSION?
S2:	YES.
I:	OK UH NOW YOU'RE A STUDENT HERE AT THE SUPSI RIGHT?
S2:	YES
I:	A AND YOU'RE STUDYING UH WHAT EXACTLY?
S2:	ENGINEERING
I:	ENGINEERING
S2:	MECHANICAL ENGINEERING
I:	SO WHAT DO YOU HAVE MORE PASSION FOR? ENGINEERING OR CHESS?
S2:	ENGINEERING
I:	YOU HAVE MORE PASSION…
S2:	YES
I:	…FOR ENGINEERING
S2:	YES ENGINEERING.
I:	REALLY?
S2:	YES, BECAUSE UH I IT'S IT'S VERY DIFFICULT TO TO SAY / BUT MY LIFE UH IS ONE PART THE JOB AND SECOND PART THE FREE TIME FAMILY AND UH OTHER / I I PUT TH' THE CHESS IN UH SECOND PART
I:	I SEE UH
S2:	MY UH FIRST PART IS UH UH UH (LAUGHS) A NORMAL A NORMAL JOB
I:	OK UH HUH
S2:	FOR EXAMPLE A MECHANICAL ENGINEERING UM UH MEDICAL ASSISTANT UH AND UH …
I:	MM HMM MM HMM
S2:	…TEACHER / DO / UH THESE UH ARE FOR ME NORMAL JOB
I:	MM HMM OK RIGHT
S2:	AND CHESS IS UH ONLY FREE TIME AND UH LOVE IN A LIFE

Speaker 3: Picture Task

S3:	YES UH HERE I I SEE THAT UM / I HAVE DOUBT ABOUT THE FIRST PICTURE / BUT ALL PEOPLE IN THIS PICTURE ARE USING COMPUTER UM / THEY ARE STUDYING SOMETHING AND UH
	FOR FOR EXAMPLE IN THE FOURTH PICTURE YOU SEE AN OLD MAN MAN WHICH IS UM READING IN A BOOK AND HAS A LAPTOP UMMM / I I THINK HE'S DOING A RESEARCH ABOUT SOME TOPIC AND UHH / THE FIRST THING I

	THOUGHT WHEN I SAW THIS PICTURE IS IT HE USES COMPUTER BECAUSE UH UMM UH FOR HIS / IT IS UH NOT SO HAPPY TO USE A COMPUTER BUT UH FOR HIS PROFESSION IT HAS BECOME IMPORTANT / THERE IS A LOT OF PEOPLE WHICH UH OLD PEOPLE WHICH HAVE THIS KIND OF UM UHH RELATIONSHIP WITH A COMPUTER / HE UMMM
	THE THIRD IT IS ALSO INTERESTING THE THIRD PICTURE WHERE YOU SEE UMM A LITTLE CHILD IN A SCHOOL AND UH / THE FIRST THING I THOUGHT WAS MY MY DAUGHTER WHICH UH IS UH FIVE YEAR OLD AND SHE IS UHHH BEGINNING USING COMPUTER / AND FOR YOUNG PEOPLE IT IS A VERY EASY IT IS UH SOMETHING WHICH IS VERY NATURAL AND UH / I THOUGHT THAT
	AND UH THE SAME UH ABOUT THE LAST PICTURE / HERE I SEE YOUNG PEOPLE UH / THIS IS A STRANGE PICTURE THE SECOND ONE BECAUSE THE PEOPLE ARE UM PERHAPS UH UM DRINKING SOMETHING / IT IS A PAUSE DURING THE WORK UM BUT THEY HAVE ALL COMPUTER AND SO I THINK THEY ARE RATHER FANATIC / THERE IS (LAUGHS) UH UH / YOU CAN SEE A SORT OF / THEY CANNOT LIVE WITHOUT COMPUTER EH
	AND THE UH MA(N) UH MAN IN THE FOURTH PHOTOGRAPH IN THE FOURTH PICTURE PERHAPS UH UH CANNOT AVOID TO USE COMPUTER / VERY DIFFERENT DIFFERENT WAY TO SEE
	THE THIRD THE THIRD ONE IS VERY INTERESTING / I THOUGHT SOMETHING ABOUT HER UH ORIENTAL ORIGIN UM / BUT I UH PERHAPS A SORT OF UH UMMM / YES SOMETHING THAT IS VERY VERY CONCERNED WITH REFLECTION WITH UM METAPHYSIC QUESTIONS / BUT HE IS SURROUNDED BY UM ARCHEOLOGICAL UH STUFF / PERHAPS IT IS A PERSON WHICH IS DOING SOME UH STUDY ABOUT UH AN ANCIENT UH CIVILISATION UH / BUT HE USE HE UH / I CANNOT SEE EXACTLY / I THINK HE IS NOT USING A COMPUTER IT IS RATHER DIFFERENT / BUT IT IS A VERY / IF I HAD TO MEET SOME OF THOSE PEOPLE TO SPEAK UH I WOULD UH UH CHOOSE THE FIRST ONE / IT IS A / SEEMS TO BE / IT SEEMS / IT SEEMS TO ME THAT UH IT IS A VERY INTERESTING PEOPLE

Speaker 3: Discussion

I:	IS THIS THE RIGHT WAY TO THINK ABOUT WORK AND LIFE? LET'S DISCUSS THIS TOGETHER.
S3:	I I THINK UH IT IS IMPOSSIBLE TO SAY UH IT IS UH UH THE RIGHT WAY TO THINK ABOUT WORK AND LIFE IN THE SENSE THAT UH HMM THIS IS A A SITUATION A REAL SITUATION FOR MANY PEOPLE / THINK THERE ARE PEOPLE UH WHICH UHHH WHO DO A STUDY UH VERY INTERESTING BUT UH UH THEY DO A UH JOB WHICH IS NOT SO INTERESTING / I THINK IT IS THE SITUATION FOR A LOT OF PEOPLE UH
I:	NOW DO DO YOU THINK THAT YOU ARE IN THAT SITUATION?
S3:	MY UH NO NO
I:	PERSONALLY (OVERLAPPING)
S3:	I THINK I'M NOT IN THIS SITUATION BECAUSE UH I WORK AS A RESEARCHER IN THE MATHEMATIC FIELD AND IT IS EXACT EXACTLY WHAT UH I WANT TO DO / BUT IT IS PARTIALLY UH CORRECT BECAUSE IF I UH HAD TO DO ONLY MY JOB UH THERE WOULD BE A LOT OF THINGS THAT UH I COULDN'T UH UH UH / I'M INTERESTED IN A LOT OF THING THAT UH UH I CANNOT DO IN MY JOB / ALWAYS ABOUT UH UH O MM O MM / NOT ALWAYS UH SOMETIMES ABOUT UH UH SCIENCE UHH / FOR EXAMPLE I'M VERY INTERESTED IN PHILOSOPHY BUT UH I DON'T WORK ON PHILOSOPHY / SO I THINK UH THE JOB ALSO FOR ME IS NOT SOMETHING THAT CAN COVER ALL MY INTERESTS
I:	NOW WOULD WOULD YOU SAY THAT YOU HAVE A PASSION FOR WHAT YOU'RE DOING NOW IN YOUR PROFESSIONAL LIFE?

S3:	YES (OVERLAPPING) YES YES I WOULD UH SAY THAT UHH
I:	NOW YOUR OTHER INTERESTS LIKE / OR OTHER THINGS THAT YOU'RE INTERESTED IN / WOULD YOU SAY THAT YOU HAVE A PASSION FOR THOSE THINGS MM / AND IF YOU HAVE A PASSION FOR THOSE THINGS COULD YOU PURSUE THOSE THINGS AS UH UH FOR A CAREER (OVERLAPPING) FOR A JOB
S3:	UH OTHER THINGS ...
I:	YEAH
S3:	… THAT I DO IN MY UH FREE TIME?
I:	YES THE THINGS THAT YOU HAVE A PASSION FOR
S3:	UH MMM
I:	I MEAN WOULD YOU LIKE TO I DUNNO ABANDON MATHEMATICS AND AND AND GO INTO UH PHILOSOPHY IF THIS IS UH A PASSION FOR YOU / COULD YOU SEE YOURSELF CHANGING YOUR JOB COMPLETELY?
S3:	UH UH YES YES
I:	AH
S3:	UH UH IN THIS SENSE UH UH UH I MM DO CON(?) SEE MYSELF AS A PERSON WHICH DO ALWAYS THE UH UH / I I DON'T SEE A BORDER UH UH A VERY STRONG BORDER BETWEEN MY MM PERSONAL INTERESTS IN THE FREE TIME AND MY JOB UH / IN THE SENSE THAT UH I TRANSFER UM MENTAL ENERGY UH KNOWLEDGE UH AMONG THOSE UH UH FIELDS UH / YES UH SO UH UH I THINK ...
I:	UH HOW HOW HOW HOW DO YOU MEAN I I
S3:	UH FOR EXAMPLE I UM / WHEN I AM INTERESTED IN UH PHILOSOPHY I THINK UH UMM ABOUT TOPICS UH UH WHICH CAN BE USEFUL ALSO IN MY RESEARCH UH / FOR EXAMPLE I THINK ABOUT WHAT IS A CONNECTION WITH BETWEEN SCIENCE THEORY AND REALITY / THIS IS SOMETHING WHICH HAS TO DO WITH PHILOSOPHY / BUT WHEN I WORK IN MY JOB I BUILD I CONSTRUCT MATHEMATICAL MODELLING MODEL OF REAL OBJECT AND SO THIS QUESTION UH IS IN MY MIND AND MY REFLECTION IN MY FREE TIME BECOME USEFUL TO UH TO TO TO GO ON UH IN MY RESEARCH
I:	HMM
S3:	SO UMM SO UH I I I I MM UM I SEE THERE ARE MANY THINGS UH THINGS WHICH ARE INTERESTING FOR ME / AND PART OF THESE OF THOSE THINGS THINGS ARE UH IN UH COVERED MY BY MY JOB / BUT WITHOUT A VERY VERY STRONG SHARP BORDER

Appendix C: Listener protocols

Listener 1

Speaker 1 Section 1 Concurrent		
01)	LC	I understand her better./
02)	LAR	(laughs)
03)	SS	She sounds more confident./
04)	SS	She's sure of herself./
05)	SL LQJ	Her English is not bad ... /
06)	LC	... I can understand her quite well./
07)	LINT	Uh huh./
08)	PI	She's uh OK she's struggling but she's gonna get it./
09)	LINT	Yes uh hm./
010)	LINT	And then... uh huh./
011)	LINT	I'm looking at the pictures too./
012)	LINT	Yeah./
013)	LINT	I think she thinks that the older people have a little more difficulty, and the younger people are uh, it's easier, 'cause they're together in a group./
014)	ST	She's uh, I said it before that she's sounds very confident, she's very sure of herself./
015)	SL LQJ	Her English is actually quite good.../
016)	ST	... and she does love people./
017)	LINT	Ah huh./
018)	LINT	She's uh feels uh very uh good with herself./
019)	LINT	Uh huh OK./
Speaker 1 Section 1 Retrospective		
020)	ST	I felt that she was uh very a very strong person.../
021)	ST	... and um uh possibly a well organised person, she seemed to.../
022)	O	... uh gosh I'm at a loss for words here, (laughs)./
023)	ST	I feel that she is just very sure of herself that's that's what I feel.../
024)	SS	... and she wasn't intimidated by the interview./
Speaker 1 Section 1 Concurrent with pausing		
025)	PE pro	She she speaks very clearly./
026)	ST	She uh she is uh I think she is quite knowledgeable.../
027)	SE	... she has uh maybe gone to university even .../
028)	ST	... she sounds really like a smart lady./
029)	ST	She might make uh a good social worker, somebody that would work with people even as a teacher./
030)	ST	She she likes to work with people I think, she likes to be with people./

031)	ST	She is uh very observant .../
032)	ST	... and again interested in people./
033)	ST	She's more of a people person, she likes to be, communicate directly with people./
034)	ST	Relationships are very important to her, working with people is is what she would really be good at./
035)	ST LQJ	Yeah, she's good, she would be very good with uh working with people./
Speaker 1 Section 1 Retrospective		
036)	ST LQJ	I think she would be great working with people, you know, she just, I don't know.../
037)	LAR	...I like her./
038)	LINT	If I see her I'll hug her some day (laughs)./
Speaker 1 Section 2 Concurrent		
039)	LINT	That's right./
040)	LINT	Yes./
041)	ST	She has a good attitude I think./
042)	LINT	That's true, yeah./
043)	LINT	You're not, that's, and you're not going to do it well.../
044)	LINT	Yeah./
045)	SS	She is very passionate about this./
046)	LINT	Uh huh./
047)	LINT LAR	Oo, ha ha../
048)	LAR	(laughs)
049)	LAR	Ah (surprise)./
050)	ST	Ah see she loves people that's for sure./
051)	SL	Her English is so good./
052)	PI	And she really can explain how she feels, very well./
053)	ST	And she certainly is is uh, she's a well-rounded person../
054)	LINT	Yeah../
Speaker 1 Section 2 Retrospective		
055)	ST	She's such a well-rounded person, I think. That's what I feel anyway um./
056)	SL LQJ	Her English is very good./
057)	PI LQJ	She can explain herself really well.../
058)	PC pres LQJ	...um she presents herself well./
Speaker 1 Section 2 Concurrent with Pausing		
059)	ST	Well she has, she has a good attitude though, about working./
060)	ST	OK, I just wanted to say that before you (the interviewer) started talking, asking her that question about “happy” I was going to say that she's seems like basically

		a very, a happy person../
061)	LINT	That's true, yeah that's very true, and you really.../
062)	ST	She has a good attitude I think about life./
063)	PI LQJ	She expresses herself really well./
064)	LINT	I wonder if when she says she likes to travel is it because she's interested in other countries like other cultures and the people, I just am wondering about that./
065)	ST	I think uh she um I think I said that before that she seems like a very well-rounded uh person./
Speaker 2 Section 1 Concurrent		
061)	LINT	Uh huh./
062)	ST	He sounds to me as a very gentle person./
063)	PE flow	And he's uh speaking slowly./
064)	LC	Easy to understand./
065)	PE flow	Thinking for his uh, the words to use./
066)	LAR	“They...” (laughs)./
067)	SL	Yes, he struggles a little bit with his English.../
068)	LC	...but it's really clear, very clear./
069)	LINT	Yes./
070)	LINT	Uh huh./
071)	LINT	Yes./
072)	LQJ / PE sty	He describes everything really well./
073)	LC	Because I can follow him with the pictures./
074)	LINT LAR PE pro	Mm Hmm, “information” (laughs), that's OK, don't worry./ (speaking to Speaker 2)
075)	LINT	Uh huh./
076)	ST	He's very observant, really observant./
077)	ST	Very sensitive man, this man, to people../
Speaker 2 Section 1 Retrospective		
078)	ST	He's uh, he sounds like a really, a very kind person.../
079)	ST	... and he's very observant, he, he pays attention.../
080)	ST	... he's more like a person that kind of sits a little bit on the outside and he's looking in, and he realises everything that's going on around him./
081)	ST	I think he's a little timid.../
082)	ST LAR	...but very nice./
083)	LAR	I really like him./
Speaker 2 Section 1 Concurrent with Pausing		
084)	LQJ ST	He uh I think well I feel that he really feels what the person, in the picture that he's looking at, is feeling, feels that, like he puts himself in the place of the person./

085)	ST	He yeah, I think that he always tries to put himself in the place of the other person, so that he understands what that person is doing or feeling./
086)	PE voc	He struggles with his words.../
087)	PI	... but he uh he really is able to express himself./
088)	LINT	Uh huh
089)	PE voc LINT	“Information” That's OK (laughs) that's OK if you don't know the word, because we know the word./
090)	ST	Ah, he's, yeah he's very observant, 'cause I wouldn't have noticed that.../
091)	LINT	... that he feels that one is in his home and the other is not./
092)	LAR LINT	Aha. That's cool./
093)	ST	He's um quite shy I think anyway.../
094)	O	... I don't know of course but I don't know because I don't see him.../
095)	ST	... but I feel that he's quite shy./
096)	SE	But he really wants to get ahead and to learn English as a language./
097)	ST	He's a very kind person./
Speaker 2 Section 2 Concurrent		
098)	LAR	Oh wow. Oh./
099)	ST	He's very shy, very timid./
0100)	SS	And um doesn't have the confidence that he would like to have./
0101)	LINT	And he likes to keep fit I think./
0102)	LINT	He, uh huh./
0103)	LINT	Ah so there you go (surprise).../
0104)	LINT	... he's more of a studious person./
0105)	LINT	But he does the the sport to keep fit./
0106)	LINT LAR	Ah, that's nice./
0107)	LINT	Ah so chess is sitting on the outside looking in and knowing what the other person is going to do. There you go./
0108)	LINT	Uh huh./
0109)	ST	Very observant./
0110)	LINT LAR	That's neat./
0111)	LINT	And when he becomes, if he's into chess he only has to communicate with one person, more or less./
0112)	LAR	Or chess? Oh! (laughs)/
0113)	LINT	He doesn't have to deal with a lot of people. I think he can work on his own more./
0114)	LINT	With his ... yeah./
0115)	ST	He's such a gentle person./
0116)	ST	But very timid./

0117)	SE	He's really well educated too./
0118)	LINT	Yeah./
0119)	LAR ST	(laughs) Oh he's a very interesting young man./
0120)	LINT	But that teaches him, chess teaches him, has taught him how to be observant./
Speaker 2 Section 2 Retrospective		
0121)	LINT	The chess has helped him to uh figure out what uh the other person's move is going to be so he can anticipate things he, he has trained himself to, or maybe it just comes naturally that he um, can anticipate what's going to happen. /
0122)	ST	He's very observant./
Speaker 2 Section 2 Concurrent with Pausing		
0123)	LINT	Yeah “for fun” (paraphrasing) He plays.../
0124)	LINT	Yeah./
0125)	LINT	His chess yeah./
0126)	O	I think that I've pretty well summed it up.../
0127)	LINT	... but again he is uh, because obviously chess has helped him to be more observant of what the next move is going to be he studies everything in his life, you know, he .../
0128)	LINT	... and I think that he feels uh more comfortable knowing what's going to happen next rather than just being a really free spirit and just going with the flow./
0129)	ST	He's so observant, he wants to, he needs to know the moves in his life./
0130)	LAR LINT	Maybe, yeah that's nice./
0131)	LINT	Ah he's at the university, engineering./
0132)	LINT	Uh huh./
0133)	LINT	Engineering, he probably enjoys engineering and those other subjects that he mentions later on, because it's a matter of figuring out step by step to get to the end result./
0134)	LAR LINT	Yeah that's nice./
0135)	ST	He's a very sensitive and uh compassionate person, I think./
0136)	LINT	With his job, with his jobs, if he gets uh, the jobs in the engineering, he would always uh end up with results and that's really important to him to be able to do step by step, to, to an answer./
0137)	LAR LINT	Oh, that's so sweet, yeah./
0138)	O	(off record: Sometimes my impressions you know though of people can be very wrong in the beginning, did you know that? Well, yes because...)./
Speaker 3 Section 1 Concurrent		
0139)	LINT	Uh huh./
0140)	SL	His English is quite good, he uh./
0141)	LC	I'm just trying to get a kind of a fix on him./
0142)	PE flow	He speaks quicker./

0143)	ST	I think his his uh mind goes very quickly…/
0144)	SL	…but his English is good./
0145)	LINT	Uh huh Yeah. He sees.../
0146)	LAR LINT	Oh how nice, he uh has a family of course.../
0147)	ST	... and now he's interested in young people I think./
0148)	PE flow	He speaks quickly./
0149)	PE voc LC	Uh, “the second one”?/
0150)	LC	I can't, I'm not understanding him really well here./
0151)	PE flow	He's rushing with his thoughts./
0152)	LINT	Uh huh./
0153)	LINT	He's, yeah./
0154)	LC	Oh, I don't understand./
0155)	LC	Oh, I'll have to hear it again./
0156)	LAR	Uh, Oh! (surprise)/
0157)	LINT	Oh, I'm just trying to follow what he's saying with the pictures./
0158)	LC	Oh OK, I have to listen to this again./
Speaker 3 Section 1 Retrospective		
0159)	LC PE flow	I I had difficulty understanding, uh comprehending what he was uh saying because he was speaking quicker and uh, I have to listen to it again now./
Speaker 3 Section 1 Concurrent with Pausing		
0160)	LC	Yes uh just having a little bit of difficulty understanding uh where he's going with this./
0161)	ST	He gives a lot of thought to what he's observing.../
0162)	LC	...but um, somehow I have difficulty understanding where he's going.../
0163)	PE flow	... because it doesn't, to me it's it's not flowing well./
0164)	LAR LINT	Oh, that's interesting./
0165)	LINT	Uh huh, it's a little child, and he, OK, we're at the little child./
0166)	LINT	I think that he is interested in um, he would, probably is very interested in computers, he uses a computer maybe a lot./
0167)	LC	And um, OK. I wonder, I'm, I'm, I still am not sure really where he's going with this./
0168)	PE flow	The other 2 were uh, they had a rhythm to their uh, to the interview./
0169)	LC	I'm have, I'm just having a little difficulty./
0170)	PE pro	His voice is very clear though, very clear.../
0171)	PE sty	... and he places his words well./
0172)	ST	Um he seems to be quite observant though and uh./
0173)	LC	Yeah I just have a little difficulty with this one, I'm sorry./

0174)	LC	I wonder if he's, I'm just not sure, yeah I'm just not sure where he's going with this./
0175)	SE	Computers maybe are, what he, he works with a computer a lot./
0176)	LC	Oh! I don't understand. I'm sorry, I just don't understand./
0177)	PE flow	And he speaks very quickly.../
0178)	PE flow	... and I'm, so, and it seems to be, the conversation, the interview is very scattered./
0179)	LINT SE	Well, I wonder if that's a hobby of his, archaeology or, or science or something./
0180)	LAR	Interesting. I just uh, I can't say any more, I'm sorry./
Speaker 3 Section 2 Concurrent		
0181)	SS	He sounds like he's not very sure./
0182)	LAR	Hm, Oh, interesting./
0183)	ST	He's not uh so, such a sensitive person he's more of an intellectual and ... /
0184)	ST	Extremely intellectual person.../
0185)	ST	...and so isn't so maybe in tune with people./
0186)	O	Sometimes it's really hard to understand intellectual people because they have.../
0187)	O	... they're thinking on a different level.../
0188)	O	... and so then it's really difficult for the average person to uh understand./
0189)	LINT	Hm./
0190)	LINT	Uh huh./
0191)	LINT	He walks a very narrow path./
0192)	SS	And I think he's focussed./
0193)	LINT	But ...yes./
0194)	PE voc	“and reality”./
0195)	SS	He stays focussed./
0196)	ST	A scientist. Hm./
0197)	LAR ST	He's a very interesting uh young person./
0198)	O	I think that uh his friends are very, are also intellectual people, read a lot./
Speaker 3 Section 2 Retrospective		
		no comment
Speaker 3 Section 2 Concurrent with Pausing		
0199)	LINT	I think, I think that he really uh, uh, is thinking really more about how his life should go./
0200)	LINT	He, he uh has made sure that he has the kind of job that he is really interested in, and he really loves it, and that makes him very happy./
0201)	LINT	He doesn't feel frustrated, I don't think he feels uh frustrated./
0202)	ST	I've said, I've said it again, I think I said in the very first part that I thought that he was maybe interested in science or something like that so, he's a very intellectual person, I think./
0203)	ST	He's interested in anything that's going to uh do with the sciences, math or things

		like that./
0204)	LINT	So if he's looking for a job, it's going to be certainly in that section./
0205)	LINT	I think, I think uh, I feel that uh, he's certainly uh, is uh focussed, and he knows where he wants to be in his life./
0206)	LINT	And he really works hard at that./
0207)	LINT	Mm hmm, I wonder if, he probably has only a very uh small group of uh, of uh friends that he enjoys really being with because they all have the same thing in common, and that really gives him a lot of joy./
0208)	LINT	Yes./
0209)	LINT	He really likes to bring everything uh together, for, he's looking for the one answer in his life./
0210)	ST	He's very, he's very curious./
0211)	ST	Oh, that's wonderful, uh yes he is uh very uh, a really uh intellectual person he uh has uh, is very focussed, he has that narrow path that he walks on because he really knows where he needs to get./

Listener 2

Speaker 1 Section 1 Concurrent		
		No comments
Speaker 1 Section 1 Retrospective		
01)	PE pro LC	First of all at the beginning because of her accent I had a little difficulty understanding some of the words but.../
02)	PC cont LC	... because of the context of the sentence I could understand what she meant, and.../
03)	PI	... I noticed as she spoke on she couldn't express herself…/
04)	PE voc	… because she was lacking vocabulary words um, and then.../
05)	O	... there's a tendency I know there's a lot with Italian speakers they leave on the "S" or they put the "S" on, they have a difficult time with plural.../
06)	PE gr	And I noticed sometimes she would say "computer"./
07)	O	They have a difficult time with plural./
08)	PE gr	"Computers" "computer" but it was mixed./
09)	O	I'd have to listen again to actually know if it was, um, but I noticed that./
010)	PE flow	She spoke very quickly.../
011)	SS	... which sounds like she's very confident in English but .../
012)	SS	... I think she either was nervous.../
013)	PE voc	... or was missing vocabulary words./
Speaker 1 Section 1 Concurrent with pausing		
014)	PE pro	OK It's just the pronunciation of "picture"./
015)	PE voc	I'm just thinking that it's, you know, "socialise people"…/
016)	LC	… I didn't understand some of the things she was saying exactly, I mean the gist of what she's saying …/

017)	PE voc	… but it wasn't “socialised people” or “people socialising” I guess is what she meant./
018)	PE gr	There, “see some old people to try to read in some books” it's totally, the grammar is totally false./
019)	LINT	Now that I'm looking at the pictures, I actually (laughs) that I'm thinking myself too, no, I'll go ahead./
020)	PE voc	She said also “trying to socialise”. It's like I said before with the vocabulary./
021)	PE voc PI	Having a lack of um, the words to, the correct words to use to express what she's thinking./
022)	PE gr	Yeah, I mean, the grammar of how to speak with, about the computer, she's, definitely doesn't know how to say, “when I work with a computer” or “when I use the computer”./
023)	ST PC	I think also um, either she doesn't have a lot of fantasy or has a hard time imagining how to answer questions./
Speaker 1 Section 1 Retrospective		
024)	LAR	It's funny 'cause, well.../
025)	PE (no sub-code)	...the first time I was listening more to how she was speaking and .../
026)	PC cont	...the second time I was listening more to what she was actually saying./
027)	PE gr	Um, otherwise I didn't want to repeat and say again, with the “s”s and.../
028)	PE voc	...the vocabulary and those things, yeah./
029)	LINT ST	Yeah, I mean she was saying, I don't think she's very positive about learning with the computer./
Speaker 1 Section 2 Concurrent		
		no comments
Speaker 1 Section 2 Retrospective		
030)	T	I was just thinking when you're being interviewed, when you're being asked questions even for me sometimes I get stuck and I don't maybe say what I really would like to say,.../
031)	T	I think sometimes having her answer the questions she, for instance “what are your passions?”, she has an idea…/
032)	O	… but it's like me right now, the words don't come out the way that you would really like to say .../
033)	O	... and you make a lot of mistakes too if it's not your language, I think.../
034)	SE	... also that um, I get the feeling that she speaks English, has spoken English a lot,.../
035)	SE	... but has probably made errors all the time, and .../
036)	SE	... hasn't had a chance to correct them, and .../
037)	SE	... they're just so ingrained in her head now that it's hard for her to um get that out./
038)	PE pro	Otherwise, no I don't, yeah, OK the same things, the pronunciation .../
039)	PE gr	... or using grammar incorrectly .../

040)	O	... but otherwise.../
Speaker 1 Section 2 Concurrent with Pausing		
041)	PE gr	Just the way she said it, “unfortunately if you don't have a work” at least that's how I understood it.../
042)	LAR LINT	I like what she says (laughs). I mean that was like a little word of wisdom, so.../
043)	LINT	Which I think is also important. Just I, what she's saying, I agree with her, I think it's a nice way to put it./
044)	PE gr LC PE pro	Yeah “I been working in the tourist”, I didn't understand what she meant, if she meant... I just didn't know if she said “tourist industry” and I didn't hear it or “with tourists” I didn't catch it./
045)	PE gr	“I can find some other jobs”, it's just grammar./
046)	PE gr	“See movies”, I don't know./
047)	PE gr	“Ride sometimes bicycle”, yeah, it's just the grammar that I'm catching now./
048)	PE gr	What's interesting is that even though you may hear some grammatical errors…/
049)	LC	… you still understand the context of what she's saying./
Speaker 2 Section 1 Concurrent		
		No comments
Speaker 2 Section 1 Retrospective		
050)	LAR	Wow that's amazing Well it's amazing.../
051)	SL	I think he's a new learner of English and .../
052)	PE gr SL	...he was speaking in the present tense because I don't think he understands how to speak in English the continuous how we say “he is reading”, but.../
053)	ST	... he's very very observant. He saw things that I didn't even see and um.../
054)	ST	he really probably if he would have the tools he would speak for an hour about these pictures (laughs)./
055)	PE voc PI	Um well also I think he was lacking vocabulary he couldn't always express himself the way he would like to./
056)	LAR PC cont PI	But I was impressed by the things he observed and was trying to say them./
Speaker 2 Section 1 Concurrent with Pausing		
057)	PE gr	It's just the choice of, it's the confusing, “I am...” “I see”or, you know, grammar./
058)	PE pro	And then there the accent, “other”, I know it's difficult, that word to say, so, he has difficulty saying that, pronouncing “other”./
059)	LAR PC cont	That is amazing to me, I didn't even see that there was a personal computer there and .../
060)	LC	I understand what he was trying to say …/
061)	PI	… even though he had a difficult time expressing himself./
062)	PE gr	“Of younger” yeah grammar, hmm./
063)	PC cont	I think this man is um, (laughs) he's always observing the computers (laughs) …/
064)	PE voc	… and he knows “computer game” (laughs) OK. Very well./

	LAR	
065)	PE voc	The “difference” uh “the different” and “difference”./
066)	PE gr SL	“Outside your home”, he sounds to me like he's really just starting to learn English, so, OK./
067)	PI	Well he says “yes” and then he doesn't answer. (laughs). Does that mean “no”?./
Speaker 2 Section 2 Concurrent		
		No comments
Speaker 2 Section 2 Retrospective		
068)	LAR	Well I mean I um what can I say I'm trying to think um I was fascinated by his …/
069)	ST	… he sounds like an interesting person but.../
070)	PI	I know he's having a difficult time trying to say what he wants to say (laughs)./
Speaker 2 Section 2 Concurrent with Pausing		
071)	PE gr	Yeah OK, “for hobby for relax”…/
072)	LC	… I understand what he's trying to say …/
073)	O	… but we would, I would say it in a different way, as far as my using English words./
074)	PI	I think he has a little difficulty explaining what he means …/
075)	PE flow	… um, “one part job, one part”, not using complete sentences.../
076)	SL	it's just his English at this point is um.../
077)	PC cont LAR	“A normal job” (laughs)./
Speaker 2 Section 2 Retrospective		
078)	LAR LINT	I guess just what he was saying too was, is interesting because he already has this impression of his life, and mechanical engineering.../
079)	O	...I know they're very precise and very, um, precision oriented so .../
080)	LINT	...I have a feeling that he has already sectionalised, I think he said “one part is hobby, one part is free time” so... /
Speaker 3 Section 1 Concurrent		
081)	PE gr	“computers”./
082)	PE gr	“computers”, yeah, the plural./
083)	LC	I don't understand what he was saying./
Speaker 3 Section 1 Retrospective		
084)	LAR	This speaker is very interesting to me because.../
085)	ST	...what he observes is just amazing and very full of fantasy, um, imagination, actually feels the pictures./
086)	PC cont	“Archaeological” background, I didn't even see those, I saw the columns, but I wouldn't have called it an “archaeological” background, but .../
087)	LAR PE voc	... it's interesting how he has a very good vocabulary and.../
088)	PI	... can express himself very well./
089)	PE gr	Their using the word computer is constantly used incorrectly as far as plural, non-

		plural, “the computer” “a computer”…/
090)	O	… It's a foreign word that's been thrown into the language and it's uh, the grammar is, I've noticed with so many people is, um, there's a problem there./
Speaker 3 Section 1 Concurrent with Pausing		
091)	PE gr	That's what I mean, “all using computer” not “computers”./
092)	PE pro	“The third thing” or “the first thing” …/
093)	LC	… I don't understand what he meant./
094)	PE pro	“The third thing I thought” OK. Hm./
095)	PE gr PI	So, the grammar, he's having a little difficulty expressing himself using the English words correctly./
096)	PE voc	“People”, he doesn't use the word “children”./
097)	PE pro LC	It's very difficult to understand if he means “in the first picture” and.../
098)	PE pro	... “her” or “him”./
099)	PE voc	Yeah, he doesn't, the word “research” doesn't come to mind to him, “doing some study”, so that's all./
0100)	LAR LINT	I think that's really interesting that he says that, “if I was to meet one of these people in the pictures”, it's, that's neat./
0101)	PE gr	“person”, yeah, grammar./
Speaker 3 Section 2 Concurrent		
		no comments
Speaker 3 Section 2 Retrospective		
0102)	LAR LINT	I don't know if I'm tired, but he was boring me. The first section was so interesting and he had so much fantasy, and now it's boring, talking about math and how he looks at his life, his math and ugh.../
Speaker 3 Section 2 Concurrent with Pausing		
0103)	PE voc	“Who do a study” I think he means “have studies a profession” or something, but.../
0104)	LC	... anyway I get the idea of what he's trying to say./
0105)	PE flow	I think his thoughts are racing and he's um, his thoughts are ahead of what he's actually been able to say… /
0106)	SS	… so, sounds nervous./
0107)	LAR PE voc	That's an interesting choice of a word, “a border”.../
0108)	LC	I understand what he means, but .../
0109)	LAR PE voc	... it's interesting that he chooses “border”./
0110)	PE flow	He's very fluent in the language otherwise./

Listener 3

Speaker 1 Section 1 Concurrent		
01)	PE pro	“To stay in our soc...” …/
02)	LC	… I don't know what she was saying./
03)	PE gr	She speaks with a lot of mistakes …/
04)	LC	… but I do understand her I think./
05)	PC cont	Maybe I understand her also because I kind of know what the point is that she's trying to make./
06)	LAR	Ha ha./
07)	LAR	Ha ha /
08)	LINT	Yeah./
09)	T	Sort of not easy to talk a lot about stuff like this I think./
010)	LAR	Hm it's funny …/
011)	PE gr	… sometimes she makes mistakes. If you really looked at the words she's kind of saying the opposite …/
012)	LC	… but because you understand what she's meaning it doesn't really interfere./
Speaker 1 Section 1 Retrospective		
013)	PC cont	Yeah what I felt was that there were some kind of non sequitur things that were kind of just really two or three times that I felt that there were comments that didn't really fit with what she was saying …/
014)	LC	… but also they didn't really confuse me too much …/
015)	PC ext	… because what she was saying that was important she repeated enough times and fleshed it out enough with examples and .../
016)	PE pro LC	... so there was just one word near the beginning that I didn't understand at all but um yeah.../
017)	PC cont LC	I pretty much understood the point she was trying to make sure./
Speaker 1 Section 1 Concurrent with pausing		
018)	PE flow	Just the little pause there.../
019)	T	... it made me think it's very difficult to answer these questions in a in an interview situation .../
020)	O	... and so I would even not focus too much on the first two minutes of the person's performance .../
021)	SS	... because they just may be nervous. That thought came to my mind./
022)	PE sty	Here you see in the inflection in her voice .../
023)	SE	...I think something that she's thought about a lot .../
024)	T	...and that's the hardest thing about these interviews you have the photograph and you're trying to get some kind of reaction from the student .../
025)	SE	...but uh if it's something that's close to home like this question “do you have a computer” “what's your relationship to computers” it's obviously that she's thought about a lot .../
026)	PE sty	...and the inflection in her voice shows that the real her is speaking out./
027)	PC	Yeah that's just a little example of how actually what she said it's partially maybe

		even kind of a contradiction of the point that she's trying to make…/
028)	PE sty	… but then she does go on to emphasise the point that she's trying to make again.../
029)	LC	if I've understood what she just said which yeah there was a little sort of cloudy moment there, but, yeah./
030)	PE sty	Yeah so you know, in the conclusion that she's reiterating what she thinks …/
031)	LC	… and it's clear enough I think./
Speaker 1 Section 2 Concurrent		
		No comments
Speaker 1 Section 2 Retrospective		
032)	LAR	So my impressions of this part of it was that it's funny.../
033)	PE pro	...the thing that struck me was that this Franca has a heavy accent …/
034)	LC	… um you pretty much can understand her and.../
035)	PE flow	… the choppiness sometimes of her sentences can be a bit distracting.../
036)	I	...but then when the interviewer was asking a question sort of trying to make it more palatable or more easily understandable or more answerable for her the question was quite truncated too.../
037)	O	... and then I was thinking well it's um yeah we there you go we maybe don't speak in such perfect sentences and questions as we think we do./
038)	LC	Yeah but she is definitely a bit special in terms of some the odd little piece that it's very hard to understand./
039)	LQJ PC	But again she seems to always sort of finish well , she makes her point./
Speaker 1 Section 2 Concurrent with Pausing		
040)	PE flow	Yeah again here her voice starts using cadences .../
041)	SE	... and she speaks naturally when it's something that she has thought about .../
042)	LAR	... and uh so that's kind of an interesting little thing .../
043)	LQJ SL	... um as far as if the quality of her English is really better I don't know but yeah.../
044)	SS	... you get some more, she's more relaxed./
045)	LC	Here again. To be quite honest I didn't really understand which word she used but you figure it out in the context.../
046)	PE pro	... did she say “important” it didn't sound like important but ... /
047)	LC	... I'm sure that's what she was trying to say anyway./
048)	PE pro LC	OK there I didn't quite understand the word “I been working in the tour so” “Toura soul” “I been working in the toura soul” I don't understand the word./
049)	PE pro	Hm Yeah “I'm working til the end of 'Tober” …/
050)	LAR LE	… Um it's funny because I hear kind of a little bit of myself in this, sometimes when I'm really really comfortable speaking with someone I start, in Italian, when I'm speaking Italian, I start to say “yes” “yeah” “yeh” - things like that …/
051)	PE pro	… and this Franca she doesn't say “or”, she says “O”.../
052)	PE flow	... her cadences are very Italian so .../
053)	O	I wonder if it kind of gets more Italian she more comfortable she is. That's just a little thought. /

Speaker 2 Section 1 Concurrent		
054)	LAR PE pro	This is a very interesting accent./
055)	ST	Hm he's observant, I didn't notice that./
Speaker 2 Section 1 Retrospective		
056)	PE pro LAR	OK so he's got quite an interesting accent I can't really pinpoint it although knowing the name Radovan, he's obviously from ex-Yugoslavia um.../
057)	SL	...his level is much different from the previous from Franca previous person um and but.../
058)	PI	...he's managing quite well./
059)	ST	He's observing things that I I didn't really pick out in these pictures./
Speaker 2 Section 1 Concurrent with Pausing		
060)	PE voc	“Search”.../
061)	O	… I always wonder if they're using that because they know the word “search” from computers, or if they know that from a dictionary, or if they've learned that “cercare” the translation is “search” anyway just a little thing./
062)	PE pro	“CCCHHHEEE” That's not an Italian “hh” sound that's “CCHH” I don't know anyway./
063)	PE pro	He seems to take care with his pronunciation or.../
064)	PE pro	could it just be that his mother tongue is closer to ours in some phonemes - “information”.../
065)	LQJ PE pro	... I find that he says that well./
066)	PE gr	“He/she/they”…/
067)	SL	… He has a lower level …/
068)	PE gr	… he's confusing the pronouns but.../
069)	LC	... it's OK./
070)	PE pr	“Informetica” That's an interesting pronunciation./
071)	PE pr	Oh you just hear “OULDER” it's quite a thing to get around “OULDER” yeah./
072)	PE pr	“ONE MAIN” it's just it's an unusual accent./
073)	PE voc ST LQJ	“Llibrary” good – good student. He didn't say “bibliotec”./
074)	PI	Does he know what the word “yes” means? He didn't say anything after that./
075)	ST	So his mother tongue is obviously different./
076)	ST	He's honest./
077)	PI LQJ	Wow. He's speaking so well./
078)	PE sty	Hear that inflection in his voice?/
079)	LINT	Hm (surprised).../
080)	LINT	Uh huh./
Speaker 2 Section 1 Retrospective		

081)	LAR / SE	It just strikes me again how comfortable people are when you ask them about things that they have thought about and things that are close to their heart um and.../
082)	LINT	... this person, he really loves chess but he really loves.../
083)	LAR	... this is interesting.../
084)	LINT	... mechanical engineering and .../
085)	LINT	... I've got lots of thoughts about that because.../
086)	LINT	... he's from ex-Yugoslavia and a job to them means a whole lot different thing than what a job means for us bored westerners they it's just you know it's like they've come from a war situation or maybe he's even lost family members and to be able to study in this country of Switzerland, does this mean he can stay here, does he have a permit to live and work here or is he just studying here I don't know these things, but I can imagine that for someone who comes from a country that has known war, a job, uh an interesting job is also a passion, it's a wonderful thing, and uh yeah so./
Speaker 2 Section 2 Concurrent		
		No comments
Speaker 2 Section 2 Concurrent with Pausing		
087)	LINT	So he said he's not good at playing football – he said it immediately./
088)	PE pro	"I don't have a talent" That's slavic "ccchhhave"./
089)	LAR LINT	Oh because "The sport it's important" - that's an interesting thing./
090)	PE pro	Even the way he laughed. In English we way "ha ha ha" and he says "cchhe cchhe" a little "cchhe"./
091)	LAR PC cont PI	Very interesting. He doesn't have the notions but he's telling us what he's really like to do. He'd love to be a professional chess player but he can't say all those words but he can get it across anyway./
092)	LINT	You know "YES, YES" We can see it's close to his heart./
093)	LINT	It's just an unusual hobby.../
094)	LAR LINT	... it's kind of neat to hear that a student likes this, has this little secret side, heh heh, opens his chess book and studies, hottest moves./
095)	PE sty LINT	Just his little "maybe" there is - it's full, it's yeah, it would be his hope I guess./
096)	LINT	Hm. He has more passion for engineering.../
097)	LAR LINT	... that is a bit suprising./
098)	LINT	But the way he says "YES, YES" To me that says a lot./
099)	LINT	Again I'm reading into it./
0100)	ST	He's Yugoslav.../
0101)	LINT	... um what does this mean for him though, this means a good job uh. Now if he's going back to Yugoslavia that would practically speaking that would be maybe like even helping to rebuild his country, or if he's here if his family has managed to you know come to Switzerland and they've got permits and, obviously he hasn't been here for years and years because he's got this heavy Slavic accent. Oh of course he's speaking English - anyway. Um if he's going to work in Switzerland

		yeah it's security, it's money, so you can just hear this, I think, uh all these things in that little "YES" the way he said it./
0102)	O	It's very difficult to say – again those are all my thoughts but I'm just imagining, yeah./
0103)	LAR LINT	Oh well it's just funny "My life is one part the job and another part the free time and the family" yeah so... /
0104)	LINT	...job is number one./
0105)	LINT	Well that's a free young person that's good – uh hum./
0106)	LINT	So it seems to me – yeah he does really love chess.../
0107)	LINT	... but he's also really happy to be studying and thinking of his future life as a mechanical engineer./
Speaker 3 Section 1 Concurrent		
0108)	LAR LINT	That's an interesting point/
0109)	PC cont	Oriental religion?/
0110)	LC PE pro	Ah that's the first picture! Thought he was referring to the third picture./
Speaker 3 Section 1 Retrospective		
0111)	LC PE pro	Well just um I never understood what was first and what was third and... /
0112)	LQJ PI	... it seems that Alberto um speaks really really well but.../
0113)	PE pro LC	... those "TH"s darn it those are hard ha ha – it's just that I didn't understand the difference./
0114)	LINT LAR	He had some very interesting observations./
Speaker 3 Section 1 Concurrent with Pausing		
0115)	PE pro LC	Aha. "I have doubt about the first picture" – I wasn't sure about that word./
0116)	LC	Yeah I see, I understand now what he meant, what he said./
0117)	PE pro	OK there "the thirst, the first, the third" - logically you say to yourself "the first thing I thought when I saw this picture" but.../
0118)	PE pro	... it's the word isn't clear, the pronunciation.../
0119)	LAR LINT	This is a very interesting observation that he's just going to say he's just going to make this observation. Its very interesting .../
0120)	PE pro	… because he does have a downward turn to his mouth.../
0121)	LAR LINT	...and what follows is a very interesting thing./
0122)	LINT	Yeah I agree, I can just imaging what he's saying is true./
0123)	PE pro	OK "third" this is "third" yeah./
0124)	PE pro LAR	"The Thirst thing I thought" heh heh./
0125)	LINT	Yeah, just the comment about for kids it's easy to learn how to use a computer./
0126)	PE pro	Oh at first you understand "pose" like it is a pose for the picture, they're posing,

		but what he means is it's a pause, it's a break./
0127)	PC cont	“computer fanatics.”/
0128)	PC cont	Hm “they can't live without computer” he says./
0129)	LINT	Yeah, hm, it sort of looks true eh?/
0130)	LQJ LINT	Good observation, very interesting observation./
0131)	PE pro	And here I am sure he said “third” but what he means is “first”./
0132)	ST PC cont	Oh it's just this guy Alberto is a real thinker - “metaphysic questions”./
0133)	LAR PE voc	Heh heh heh – “archeological uh stuff”./
0134)	LINT	Um hm. It's unclear as to if he's also got a computer, yeah./
0135)	PC cont	“I would choose this first picture” yeah he's uh to.../
Speaker 3 Section 2 Concurrent		
0136)	PE pro	Alberto has also quite a like a “CCHH”/
0137)	PE pro	You see that he puts effort into his pronunciation and.../
0138)	PE pro	... he's confused about the “TH” thing, it sort of runs away and kind of catches up in the most appropriate moments./
0139)	PE sty	He's the opposite of Franca eh. He takes such care./
0140)	SE	He's thought about this./
0141)	LC	Ah. That was unclear./
0142)	LINT	Lucky. Heh heh./
0143)	LAR LINT	Uh huh. That's very interesting./
0144)	SE	This is all stuff he's thought about…/
0145)	PE flow PI	… so he's very fluently explaining himself./
Speaker 3 Section 2 Retrospective		
0146)	PE voc	What has just come into my mind is you don't hear him saying “first” or “third” any more.../
0147)	O	and it makes me think that when someone is speaking freely about subjects that they're comfortable with I think we all choose our patterns of speech, uh, what am I trying to say, if you're not comfortable with certain expressions for example in this case the ordinal numbers you don't go there in your normal speech 'cause you're not sure ha ha ha and uh but you know when you're constrained to answer a question uh you're trying to be academic and you're trying to fit into this kind of, “the first second and third picture” and uh, well maybe it's kind of obvious that you have to use those terms.../
0148)	PE pro	... but, you just don't hear him making those pronunciation errors uh .../
0149)	PC cont	... when he's freely talking about things that interest him.../
0150)	PE flow	... he flows, he flows really well./
Speaker 3 Section 2 Concurrent with Pausing		
0151)	LAR T	I like this right away, the guy's a total, He right away takes issue with the

		question./
0152)	ST	He's an intellectual.../
0153)	SE	... he's very self assured, and.../
0154)	LINT	... he knows that these questions that you can do anything with these questions really and uh whereas.../
0155)	ST	... maybe someone who is not so scholastic or sure of themselves maybe that's all.../
0156)	SL	...he's just able to kind of say what he wants to say and you know.../
0157)	LAR LINT	... “I think it's impossible to answer this question” is basically yeah it's kind of interesting that he does that./
0158)	LC PC cont	This was confusing to me: “I think there are people who do a study which is very interesting to them”./
0159)	LINT	I think what he means to say is “I think there are people who study and then they have their work”./
0160)	-	Anyway continue./
0161)	LINT	He means that people study and work, and their study is where their real interest lies their course of study or whatever, their night's classes./
0162)	LINT	I was just thinking it's a big kind of thinker that can... he's fulfilled in his work and, but he sees that there are other people who are not as fulfilled as he is, and who suffer in their.../
0163)	LINT	...maybe he sees this also from, maybe his father was never fulfilled in his job and counselled his son to just follow his most, his strongest interests.../
0164)	LAR	...and I think that's interesting./
0165)	PE flow sc	Oh he's just able to correct himself and backtrack and choose the right word and.../
0166)	SS	... I think he's sure of himself./
0167)	LAR LINT	Oh it's just wonderful that he can say he has a passion for what he's doing in his job./
0168)	PE sty	And he says it unreservedly./
0169)	SS LINT	This is, he's very, well I was going to say he's very self assured uh because you know he's happy with his job, he's a mathematician and uh.../
0170)	LINT	...it's his passion and .../
0171)	LINT	...he knows that not everybody can say that and yet.../
0172)	LINT	...he says he could see himself “abandon” was the word the interviewer used, he could see himself just leaving it all and choosing another field./
0173)	LAR	Well I find that laudable and yet .../
0174)	LINT	I just though to myself after, does he really know what he's saying? I mean because maybe the fact that he works in something, what if he didn't find a job in philosophy would he feel that passion, as it were.../
0175)	LAR LINT	This is nice. That's nice./
0176)	PC ext	So things kind of overflow and uh.../
0177)	ST	... he's very open minded, he doesn't compartmentalise, and .../

0178)	LINT	...what he learns in, he tries to apply things to uh what he's doing./
0179)	LAR	It's a teeny bit abstract for me.../
0180)	LS	... and I'm tired now, but anyway./

Listener 4

Speaker 1 Section 1 Concurrent		
01)	PE pro	OK she's got a very kind of slushy mixed together pronunciation it seems./
02)	PE pro	Sounds Spanish to me./
03)	PE pro	OK she's expressing herself with fairly clear pronunciation.../
04)	PC cont	... but some difficulty defining her ideas./
05)	PE flow	OK she's speaking kind of fluently.../
06)	PE gr	...and uh and clearly making some grammatical mistakes .../
07)	PI	but communicating well./
08)	PI	She's responding well to the questions that are asked./
09)	PE flow	She's expressing herself with a certain amount of awkwardness.../
010)	PI	but nevertheless she's getting her idea across, seems to me./
Speaker 1 Section 1 Retrospective		
011)	ST	Well she seems um she seems willing to speak.../
012)	PI	and to communicate.../
013)	PE flow	... um she's she's in a sense fluent because she's speaking you know kind of fluently.../
014)	PE flow	... and uh stumbling along .../
015)	PE gr	... um making lots of mistakes .../
016)	PI	... um but managing to express what she wants to say.../
017)	PC ext	...although it seems she's also having difficulty coming up with sort of ideas of any depth to express in talking about these pictures./
Speaker 1 Section 1 Concurrent with pausing		
018)	SS	OK in starting off maybe it's because she's a little bit anxious .../
019)	SS	... I think it improves a bit .../
020)	PE flow	... but I think I said the first time it sounds kind of slushy to me that she's um really stumbling over herself as she tries to go along there .../
021)	PE gr	... um making several mistakes um grammatically .../
022)	PE pro	... um a very strong accent .../
023)	LC	... but nevertheless um easy enough to understand, to comprehend./
024)	PE flow	OK lots of hesitation there.../
025)	PE flow	... some repetition .../
026)	PC	... and um finally kind of really fizzling out and not able to extend it any more than

		that./
027)	PE gr	Grammatically she's leaving out the verb “be” in the present continuous there it seems um .../
028)	LC	... Uh still perfectly comprehensible .../
029)	LE	... and it's something that's not not unfamiliar to me as a teacher to hear so I I can follow it .../
030)	O	... I think uh um people who are not English teachers might have a little more difficulty, but not much, following what she's saying .../
031)	PE flow	This whole section is a lot more fluent .../
032)	LC	... and comprehensible .../
033)	PC ext	... and a little deeper that what has gone before maybe./
034)	SS	My my impression is that maybe she was a little bit anxious at first .../
035)	PE flow	... and uh now she's she's slowed down kind of collected her her thoughts .../
036)	LC	and is expressing herself a little more clearly./
037)	PE sty	She's using a lot of good intonation also .../
038)	PE sty	... um emphasising key words and um in that way her .../
039)	PI	... uh her communication has improved from the beginning of this part./
Speaker 1 Section 2 Concurrent		
040)	I	This is a long, where you're talking (laughs) .../
041)	O	... now she's responding./
042)	PI LQJ	And she's responding well to the question and uh the input of the examiner./
043)	PE gr	Again sprinkled with grammatical mistakes .../
044)	LC	... but nothing that's really getting in the way of understanding her communication./
045)	PE sty	She's expressing herself with uh a certain amount of uh force and um emotion again the intonation is quite good, expressive./
046)	PI	There have been several questions by the examiner all of which she understands and responds to immediately./
047)	LINT	OK she's talking about her free time activities and kind of listing them./
Speaker 1 Section 2 Retrospective		
		No comments
Speaker 1 Section 2 Concurrent with Pausing		
048)	PI	OK she begins with uh some long sentences where she she really gets into the question here .../
049)	PI	... and expresses some ideas and .../
050)	PE conn	... and connects ideas .../
051)	PC cont	... and talks about the consequences of certain actions .../
052)	PC ext	... so it's fairly uh complex .../
053)	LC	... and clearly, clearly stated .../
054)	PE gr	... despite a few grammatical mistakes./

055)	PI LQJ	OK she responds very well to the long question the examiner asks (laughs) .../
056)	PC cont	... um and again makes valid clear points./
057)	PE gr	OK in responding to this question and speaking grammatically she uses some different verb tenses um usually successfully like with the present perfect continuous, but then the present perfect simple she doesn't use the correct past participle, however she mixes up verb tenses very well in this part I think./
Speaker 1 Section 2 Retrospective		
058)	PI	She responds very well to all of the questions that are put to her .../
059)	PC ext	... um not always at great depth .../
060)	PI	... but she immediately understands .../
061)	ST	... and is very willing to respond and to communicate .../
062)	LQJ	... and does it pretty well./
Speaker 2 Section 1 Concurrent		
063)	SL	OK this is pretty obviously a speaker at a fairly low level of ability in English./
064)	PE gr	OK making very basic errors .../
065)	PE flow	... speaking very slowly and carefully./
066)	PE voc	Vocabulary errors .../
067)	PE gr	... grammatical errors./
068)	PE voc	Again vocabulary errors./
069)	PE flow	He's repeating the same patterns over and over again../
070)	PE gr	Only one possessive adjective is used, “your”, it's always “your”./
071)	PI	He says “yes” but means “no” apparently (laughs)./
Speaker 2 Section 1 Retrospective		
		No comments
Speaker 2 Section 1 Concurrent with Pausing		
072)	PI	OK he's expressing himself in just as simple a way as he possibly can .../
073)	PE gr	... using the most basic forms .../
074)	PE gr	... and uh not even a third person “s”./
075)	PE flow	He's speaking very slowly .../
076)	PE flow	... um searching for almost every word .../
077)	PE flow	... um being, trying to be very careful I think to form his sentences .../
078)	PC cont	... and say what he wants to say .../
079)	PC ext	... but in as simple a way as possible./
080)	PE flow	Um because of the way he's speaking every word is completely separate there's no fluency at all it's almost like a machine producing a list of words .../
081)	PE flow sc	... um he does correct himself at one point he corrects the pronoun um./
Speaker 2 Section 1 Retrospective		
082)	PI	Apparently he didn't understand the final question “is there anything else you want to say?” because he answers yes but then says nothing until the examiner ends that

		section./
083)	LAR PE flow	So it's almost painful to listen to, the words coming out drop by painful drop (laughs)./
Speaker 2 Section 2 Concurrent		
084)	PI	First he responds just with yes or no .../
085)	PC ext	... then the third question he goes to some length ./
086)	PE flow	Still very very slow, word by word./
087)	PI PC ext	He's responding to questions with one word answers and then the examiner is stimulating him to go a little further./
088)	PE gr	I don't think he's used any verb tenses except the present simple .../
089)	PE voc	... and very limited vocabulary./
090)	PE voc	Searching for words./
Speaker 2 Section 2 Retrospective		
		No comments
Speaker 2 Section 2 Concurrent with Pausing		
091)	PC ext	OK he's um he's choosing the easiest possible way I think to to communicate to express himself and he's really limiting his expression so he doesn't get into any trouble .../
092)	PE sty	... he's speaking extremely carefully as I said before .../
093)	PE flow	... word by word .../
094)	SL	... which at his level is probably not a bad idea./
095)	O	I I think he would have a lot of difficulty communicating with people in normal situations because most people would not have the patience to listen long enough to hear even one sentence./
Speaker 2 Section 2 Retrospective		
096)	PC ext	'K that's the end and uh although he speaks at some length in in time in talking about um chess versus engineering .../
097)	PE flow	... um it takes him a very long time to express a very simple idea because of the, the plodding way that he proceeds in putting his sentences together .../
098)	PE sty	... there's there's very little emotional content it seems as I think I said before it sounds almost like a computer generated word by word speaking. There's not much intonation uh expression beyond the meanings of the words .../
099)	PE gr	... and uh nothing but the present simple .../
Speaker 3 Section 1 Concurrent		
0100)	PE flow	I have an immediate impression that uh he's going to be rather fluent.../
0101)	PC ext	...and have something to say .../
0102)	PE flow	...he's jumping right in .../
0103)	PE flow	...and moving along./
0104)	PE voc	He's doing – using some good vocabulary “laptop” “research” “topics”./
0105)	PE gr	Some minor errors .../
0106)	PE gr	...but a good mix of verb tenses usually used correctly./

0107)	PE gr	Several small errors .../
0108)	LC	... but nothing that's getting in the way of communication./
0109)	PC	And he's saying quite a lot about each of the pictures./
0110)	PC ext	He's very able to go on and continue and extend./
0111)	PE voc	Some more very good vocabulary.../
0112)	PC ext	... needs absolutely no stimulus from the examiner./
0113)	PE sty	Good intonation good expressiveness in his voice./
0114)	PE voc	Some more very good vocabulary being used./
0115)	PE gr	Using the second conditional there correctly./
0116)	SL	So he seems to have a good level of ability./
Speaker 3 Section 1 Retrospective		
		No comments
Speaker 3 Section 1 Concurrent with Pausing		
0117)	PC cont	He immediately grasps um the similarity of the pictures the common theme that the pictures have and begins to talk about that./
0118)	PC ext	OK he's also going beyond simple description of the picture .../
0119)	PC cont	... and speculating about the people's feelings or motives for what they're doing./
0120)	PC ext	Again he's he's going beyond mere description and um talking about his personal reaction and his personal experiences uh that are related to these pictures and this topic./
0121)	ST	He's he's he exhibits a lot of uh maturity as well as linguistic ability .../
0122)	ST	... and intelligence I would say, he seems to be an intelligent person./
0123)	PC ext	Again he's speculating about the people, not merely describing what he sees but uh talking about how they may feel um and why they're doing what they're doing./
0124)	LQJ PC cont	Again he's speculating about the people in the photographs, he compares 2 of them very well./
0125)	ST PC cont	and um he seems to have a lot of imagination as well he can uh invent, if you like, different motives that people have, different feelings that they're experiencing./
0126)	PE gr	Hm. And the second conditional that he uses correctly .../
0127)	LAR LINT	... he's imagining meeting one of these people and he chooses the person he would like to meet if he had to meet one of them, which is interesting./
Speaker 3 Section 1 Retrospective		
0128)	LAR LINT	So he's uh very interesting to listen to .../
0129)	PC ext PI	... he goes into a lot of depth and imagination and expresses himself very well .../
0130)	PC ext	Extends enormously well, the examiner doesn't have to say a word except right at the beginning “What do you think about these pictures?” and he's off.../
Speaker 3 Section 2 Concurrent		
0131)	PI	Again he jumps right into the topic./
0132)	PE flow	He's speaking very very easily and fluently./

0133)	ST	Although he's obviously not a native speaker .../
0134)	PE flow	... he expresses himself in a very natural way uh pausing, emphasising.../
0135)	I	There's a long question by the examiner here, (laughs) who expresses himself fairly clearly.(laughs)/
0136)	PC ext	When asked to clarify a point he gives a good example .../
0137)	LC	... and is successful in communicating what he wants to say it seems./
Speaker 3 Section 2 Retrospective		
0138)	PE flow	He goes on in a very natural way./
0139)	PE gr	He's got some some errors that um like using “which” to refer to people um some kind of basic things.../
0140)	PE voc	... but also a pretty wide vocabulary.../
0141)	PE flow	... and uh a very natural way of speaking .../
0142)	ST	... um as I said before he seems to be very intelligent.../
0143)	PC ext	and uh have a lot of things to say, he can expand as long, as long as you want him to, it seems/
Speaker 3 Section 2 Concurrent with Pausing		
0144)	PC cont	OK he immediately puts the topic question into a real context, talking about real people's situations .../
0145)	LQJ LINT	... and makes a good point there./
0146)	PC ext	Again he goes on at some length.../
0147)	PC ext	... and giving examples and clarifying .../
0148)	PC ext	... and extending what he wants to say um. /
0149)	PE flow	He also has a kind of natural rhythm that I think of as being uh, very British English.../
0150)	O	... where people will speak very slowly and then they'll put a whole bunch O' words together real fast. This to me is a typical educated British way of speaking./ ...
0151)	SE	... and he seems to have learned that./
0152)	I	The examiner has to really clarify what he's, the information he's trying to get from uh from this guy./
0153)	PI	and uh then he is a little bit slow getting into this response it seems./
0154)	PI	OK, now he seems to be plugged into the idea, the question there.../
0155)	PC ext	and uh gives some good clear examples of what he's trying to say, some concrete examples./
Speaker 3 Section 2 Retrospective		
0156)	ST	So he's uh very willing, very willing to speak .../
0157)	PC ext	... and to expand on what he wants to say, give examples.../
0158)	PE voc	... he's got a good vocabulary.../
0159)	PE flow	... and you know, speaks very fluently./

Appendix D: Listener verbalisation codes

In order to avoid the need to constantly refer back to this list, Roman numerals that are traditionally used for sub-codes (PE i, PE ii...) in coding systems have been replaced with a few lower-case letters.

Listener verbalisations of thoughts referring to something about:

S	**THE SPEAKER**

ST	**Speaker Traits -** personality / work ethic / willingness to speak / observant
SS	**Emotional State -** Comfort level / relaxed / nervous
SE	**Personal Experience**
SL	**Linguistic Ability Level –** quality of English

P	**THE SPEAKER'S PERFORMANCE**

PC	**Content**
PC cont	Simply noting the presence of an idea / a point being made / what is being said / topic / fitting context / content fitting together/ reporting content without contributing to it – see LINT (below) / direct quoting
PC ext	Extension / development of ideas / simplicity / complexity / depth / giving examples
PC pres	Self presentation

PE	**Manner or Form of Expression**
PE conn	**Ideas Connected /** connecting of ideas
PE flow	**Flow /** fluency **/** flow / rhythm / cadence / awkwardness / repetition / hesitation
PE flow sc	**Self Correction**
PE gr	**Form** / structure / grammar
PE pro	**Sound Formation /** manner of pronunciation / accent L1 interference
PE sty	**Style** - with force / emotion / expressiveness / emphasis / carefully / inflection / descriptive
PE voc	**Vocabulary** – words / appropriacy

PI	**Interaction and Communication**
PI	Responding to questions / comprehension of task / question / references to communication / explaining self / expressing self / struggling to communicate

L	**THE LISTENER or SOMETHING LOCATED WITHIN THE LISTENER**

LAR	**Listener Affective Reaction**
LC	**Listener Comprehension –** degree of comprehension in listener
LE	**Listener Experience**
LINT	**Listener act of mimicking "interaction" with speaker: by paraphrasing, responding to or interpreting content / noting content but adding some listener contribution**
LQJ	**Listener Qualitative Judgement**

T	**THE TASK**

I	**THE INTERVIEWER**

O	**OTHER REFLECTIONS / SPECULATIONS**

	NS qualities in general
	NS speech in general
	NNS qualities in general
	NNS L1
	Teacher's versus non-teachers

Appendix E: Listener protocol data

Code	Li1	%	Li2	%	Li3	%	Li4	%
ST	58	26.6	5	4.0	10	4.7	8	4.9
SS	8	3.7	3	2.4	1	1.9	3	1.8
SE	5	2.3	4	3.2	7	3.3	1	0.6
SL	7	3.2	4	3.3	4	1.9	3	1.8
∑ S	**78**	**35.8**	**16**	**12.9**	**25**	**11.8**	**15**	**9.1**
PC cont	0	0	7	5.6	11	5.2	9	5.5
PC ext	0	0	0	0	2	0.9	24	14.6
PC pres	1	0.5	0	0	0	0	0	0
∑ PC	**1**	**0.5**	**7**	**5.6**	**13**	**6.1**	**37**	**22.6**
PE conn	0	0	0	0	0	0	1	0.6
PE flow	10	4.6	4	3.2	7	3.3	28	17.1
PE flow sc	0	0	0	0	1	0	1	0.6
PE gr	0	0	23	18.5	4	0.5	18	11.0
PE pro	3	1.4	9	7.3	30	14.2	4	2.4
PE sty	2	0.9	0	0	8	3.8	6	3.7
PE voc	4	1.8	16	12.9	4	1.9	9	5.5
∑ PE	**19**	**8.8**	**52**	**41.9**	**53**	**25**	**67**	**40.9**
PI	**6**	**2.8**	**3**	**2.4**	**6**	**2.8**	**21**	**12,8**
∑ ALL P	**26**	**11.9**	**62**	**50.0**	**72**	**34.0**	**124**	**75.6**
LAR	24	11.0	15	12.1	29	13.7	3	1.8
LC	19	8.7	12	9.7	21	9.9	8	4.9
LE	0	0	0	0	1	0.5	1	0.6
LINT	53	24.3	8	6.5	46	21.7	1	0.6
LQJ	10	4.6	0	0	7	3.3	5	3.0
∑ L	**106**	**48.6**	**35**	**28.2**	**104**	**49.1**	**18**	**11.0**
T	**0**	**0**	**2**	**1.6**	**4**	**1.9**	**0**	**0**
I	**0**	**0**	**0**	**0**	**1**	**0.5**	**3**	**1.8**
O	**8**	**3.7**	**9**	**7.3**	**6**	**2.8**	**4**	**2.4**
TOTAL	**218**		**124**		**212**		**164**	

While the total number of identifiable and usable codes from the 4 listeners is 718, the number of protocol lines was 666. The difference in the totals is accounted for in two ways. Utterances such as

L1 09 Yes uh hm./
L1 010 And then ... uh huh./

seem to occur when the listeners mimic interaction in some way with the unseen speaker, as though they were conversing. They were therefore coded as LINT, but owing to the shortness of the protocol and other possible interpretations, this coding was deemed unreliable. These codings were not counted. This type of protocol occurred more in L1 than other listeners (29 times). Further, a number of protocols received more than one coding depending on how the utterances were expressed. If 2 thought processes in a single sentence were said clearly one after the other, they were split. If this was not so clear especially if the thoughts appeared to be intertwined, they were left as one protocol, and labelled with 2 codes.
For example the utterance:

... um making several mistakes um grammatically um a very strong accent, but nevertheless um easy enough to understand, to comprehend.

allowed for straightforward coding:

L4 021 ...um making several mistakes um grammatically.../
L4 022 ...um a very strong accent.../
L4 023 ...but nevertheless um easy enough to understand, to comprehend./

Occasionally, splitting would make utterances more difficult as in:

L3 121 and what follows is a very interesting thing./

Here, separating “very interesting” referring to the LAR, from “what follows is ... a thing”, the object of the LAR which was coded as PC, was deemed unnecessary. While this is done to assist the reader, the data below focusses on coded protocols, and so the multiple coding of individual utterances will not affect our quantitative data..

Although this did not always contribute to answering our research questions and therefore not extensively discussed in the body of this paper, it may be interesting to note certain pairings of codings. For example, LAR (Listener Affective Reaction) or LQJ (Qualitative Judgement) was normally paired with a specific cause of the affective reaction or judgement. These included Speaker content, pronunciation, grammatical error or even the Interviewer's input. Other pairings occurred often with LINT and LC.

Bibliography

Bachman, L.F. 1990. *Fundamental Considerations in Language Testing.* Oxford: OUP.

Bachman, L.F. 2005. "Building and Supporting a Case for Test Use." *Language Assessment Quarterly*. 2/1, 1-34.

Bachman, L.F. & Palmer, A.S. 1996. *Language testing in Practice*. Oxford: OUP.

Banerjee J. & Luoma, S. 1997. Qualitative Approaches to Test Validation. In Clapham, C & Corson, D. eds. 1997.

Barnwell, D. 1989. "'Naive native speakers and judgements of oral proficiency in Spanish" *Language Testing*, 6/2.

Bernstein, J. 1997. "Speech recognition in language testing" In Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Brooks, L. 2003. "Converting an Observation Checklist for use with the IELTS Speaking Test." Research Notes. 11, 20-21.

Brown, A. 2003. "Interviewer variation and the co-construction of speaking proficiency" *Language Testing*, 20/1, 1-25.

Brown, A., Iwashita, N., McNamara, T. and O'Hagan, S. 2001. Investigating Raters' Orientation in Specific-purpose Task-based Oral Assessment. Paper presented at the 23rd Annual Language Testing Research Colloquium, St. Louis, Missouri. Feb. 20-24, 2001. Summary available at http://www.iltaonline.com/ILTA_archive/LTRC23.pdf

Brown, A., Iwashita, N. and McNamara, T. 2005. "An Examination of Rater Orientations and Test-Taker Performance on English-for-Academic-Purposes Speaking tests." *Monograph Series.* 29.

Cambridge University Press, Online catalogue 2006, available at http://www.cambridge.org/elt/catalogue/adult/, accessed on 01.04.2006.

Canale M., Swain M. (1980), "Theoretical bases of communicative approaches to second language teaching and testing", *Applied Linguistics*, I,1,1-47.

Clapham, C & Corson, D. eds. 1997. *Language testing and assessment. Encyclopedia of language and education v. 7.* Dordrecht: Kluwer.

Council of Europe, 2001. Common European Framework of Reference For Languages (CEFRL) Electronic Edition. Available at http://culture2.coe.int/portfolio/ (Accessed 15.12.2004)

Council of Europe, 2003. *Relating language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment (CEF). Manual: Preliminary Pilot Version*. DGIV/EDU/LANG 2003, 5. Strasbourg: Language Policy Division. Available at http://culture2.coe.int/portfolio/ Accessed 15.03.2006)

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T. and McNamara, T. 1999. *Dictionary of language testing.* Cambridge: CUP

Douglas, D. 1997a. Language for Specific Purposes Testing. In Clapham, C & Corson, D. eds. 1997 *Language testing and assessment. Encyclopedia of language and education v. 7.* Dordrecht: Kluwer. 111-119.

Douglas, D. 1997b. "Testing Speaking ability in Academic contexts: Theoretical Considerations." TOEFL Monograph Series, 8. Princeton, New Jersey: ETS.

Douglas, D. 2001. "Language for Specific Purposes assessment criteria: where do they come from?" Language Testing. 18/2. 171-185.

Erdosy, M. U. 2000. Exploring the Establishment of Scoring Criteria for Writing Ability in a Second Language: The Influence of Background Factors on Variability in the Decision-making Processes of Four Experienced Raters of ESL Compositions. Unpublished MA thesis U of T.

Ffrench, A. 2003. "The development of a set of assessment criteria for Speaking Tests". *Research Notes* 13, 8-16. University of Cambridge ESOL Examinations.

Foulkes, I. 2003. "English powers ahead". Swissinfo Swiss Radio International. Available at http://www.swissinfo.org/eng/Swissinfo.html?siteSect=105&sid=1329636 (Accessed on 11.05.2006)

Fox, J. 2004. "Biasing for the Best in Language Testing and Learning: An Interview with Merrill Swain." *Language Assessment Quarterly*. 1/4, 235-251.

Fulcher, G. 1987. "Tests of oral performance: the need for data-based criteria." *ELTJournal* 41, 287-291.

Fulcher, G. 1996a. "Does thick description lead to smart tests? A data-based approach to rating scale construction" *Language Testing* 13, 2, 208 - 238.

Fulcher, G. 1996b. "Testing Tasks: Issues in Task Design and the Group Oral." *Language Testing* 13, 1, 23 - 51.

Fulcher, G. 2003. *Testing Second Language Speaking*. Harlow: Longman.

Fulcher, G. 2004. "Are Europe's tests being built on an 'unsafe' framework?" Available at http://education.guardian.co.uk/tefl/story/0,5500,1170569,00.html. Accessed on 02.01.2005.

Glavinic M.L. 2005. Developing an online Community of practice: a case study of professional development needs for ESL practitioners working with the Canadian Language Benchmarks. MA thesis U of T.

Green, A. 1998. *Verbal protocol analysis in language testing*. CUP: Cambridge.

Hafernik, J.J. Messerschmitt, D.S. & Vandrick S. 2002. Ethical Issues for ESL Faculty. New Jersey: Erlaub.

Hamp-Lyons, L 1997: "Ethics in Language Testing" In Clapham, C & Corson, D. eds. 1997. *Language testing and assessment. Encyclopedia of language and education v. 7.* Dordrecht: Kluwer. 323-333

Hill, K. 1997. "Who should be the judge? The use of non-native speakers as raters on a test of English as an International Language". In Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Hubbard, C., Gilbert, S and Pidcock J. 2006 in press. "Assessment processes in Speaking tests: a pilot verbal protocol study. *Research Notes*.

Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.) 1997. *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Jacoby, S. & McNamara, T. 1999. "Locating Competence" *English for Specific Purposes*. 18/3, 213-241.

Jakobovits, L.A. 1972. "Authenticity in foreign language teaching. Introduction to Savignon, S. *Communicative competence: an experiment in foreign language teaching*. Available at http://www.soc.hawaii.edu/leonj/leonj/leonpsy/instructor/cc.html

Jones, N. 2005. "Raising the Languages Ladder: constructing a new framework for accrediting foreign language skills." *Research Notes*. 19, 15-19.

Knight, B. 1992. "Assessing speaking skills: a workshop for teacher development." *ELT Journal*. 46/3, 294-302.

Lantolf, James P., and William Frawley. "Oral Proficiency Testing: A Critical Analysis." *Modern Language Journal* 69 (1985): 337 – 345.

Lantolf, J. P. & W. Frawley. (1988). "Proficiency: Understanding the construct." *Studies in Second Language Acquisition*, 10: 181-196

Lazaraton, A. 1996. Interlocutor support in oral proficiency interviews: the case of CASE. Language Testing, 13/2, 151-172.

Lazaraton, A. "Evaluative Criteria for Qualitative Research in Applied Linguistics: Whose Criteria and Whose Research?" *The Modern Language Journal* 2003 87/1, 1.

Lazaraton, A. & Wagner, S. 1996. "The Revised Test of Spoken English (TSE): Discourse Analysis of Native Speaker and Nonnative Speaker Data." TOEFL Monograph Series. 7. ETS.

Kunnan, A.J. (Ed.) 2000. *Fairness and validation in language assessment*. Cambridge: Cambridge University Press.

Maldonado, C. S. 1994. "Some Aspects of "Foreignness" in the Pronunciation of Upper Intermediate English Students of Spanish." *Edinburgh Working Papers in Applied Linguistics*. 5, 78-97.

Mathew, R. 2004. "Stakeholder Involvement in Language Assessment: Does it Improve Ethicality?" *Language Assessment Quarterly*. 1/(2&3), 123-135.

McKinley, M.L. 2004 Ph.D. Thesis Large Scale Performance-Based assessment: Identification of Individual Student Gaps with Implications for Teacher Content Knowledge. U of Toronto

McNamara, T. 1990. "Item Response Theory and the validation of an ESP test for health professionals." *Language Testing*, 7/1, 52-75.

McNamara, T. 1996. *Measuring Language Performance*. Essex: Longman.

McNamara, T. 1997. "Performance Testing." In Clapham, C & Corson, D. eds. *Language testing and assessment. Encyclopedia of language and education v. 7.* Dordrecht: Kluwer. 131-139.

McNamara, T. 2001. "Language assessment as social practice: challenges for research." Language Testing. 18/4, 333-349.

McNamara, T. 2003. "Looking back, looking forward: Rethinking Bachman." *Language Testing*, 20, 466-473.

Meiron, B. E., & Schick, L. S. 2000. "Ratings, raters and test performance: An exploratory study." In Kunnan, A.J. (Ed.), *Fairness and validation in language assessment* 153-174. Cambridge: Cambridge University Press.

Messick, S. 1996. "Validity and washback in language testing." *Language Testing*, 13/3, 241-256.

Milanovic, M. & Weir, C. Eds. 2003. *Balancing Continuity and Innovation. A History of the CPE Examination 1913-2002*. Cambridge: CUP.

North, B. 1997. "The Development of a Common Framework Scale of Descriptors of Language Proficiency Based on a Theory of Measurement" in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

North, B. 2000. *The Development of a Common Framework Scale of Descriptors of Language Proficiency Based on a Theory of Measurement*. New York: Lang.

North, B. 2004. "Europe's framework promotes language discussion, not directives." Available at http://education.guardian.co.uk/tefl/story/0,5500,1191130,00.html Accessed on 01.02.2006.

North, B. & Schneider, G. 1998. "Scaling descriptors for language proficiency scales." *Language Testing*. 15/2, 217-263.

O'Malley J. M., & Valdez Pierce, L. 1996 *Authentic Assessment for English Language Learners.* USA: Addison-Wesley

Orr, M. 2002. "The FCE Speaking test: using rater reports to help interpret test scores." *System*. 30, 143-154.

Oxford University Press, Online catalogue 2006. Available at http://www.oup.com/elt/catalogue
Accessed on 04.04.2006.

Patri, M. 2002. "The influence of peer feedback on self and peer-assessment of oral skills." *Language Testing*. 19/2, 109-131.

Pollitt, A & Murray, N.L. 1996 "What raters really pay attention to". In: Milanovic N. & Saville N. (eds): *Language Testing 3 - Performance, Testing, Cognition and Assessment.* Cambridge: CUP, 74-91.

Pollitt, A. 2004. *Let's stop marking exams*. UCLES Research and Evaluation Division.

Powers, D.E, Schedl & M.A. Leung, S. W. 1999. "Validating the revised Test of Spoken English against a criterion of communicative success" Language Testing. 16/4, 399-425.

Powers E. & Stansfield C.W. 1985. "Testing the Oral English Proficiency of Foreign Nursing Graduates." *The ESP Journal*. 4, 21-35.

Rea-Dickins, P. 2004. "Understanding teachers as agents of assessment." *Language Testing*. 21/3. 249-258.

Sakyi, A. A. 2003. A Study of the Holistic Scoring Behaviours of Experienced and Novice ESL Instructors. PhD. Thesis U of T.

Salaberry, R. 2000. "Revising the revised format of the ACTFL Oral Proficiency Interview." Language Testing. 17/3, 289-310.

Savignon, S. 1972. *Communicative competence: an experiment in foreign language teaching*. Philadelphia: Center for Curriculum Development.

Saville, N. 2003. "The process of test development and revision within UCLES EFL" In Milanovic, M. & Weir, C.

Spolsky, B. 1985. "The limits of authenticity in language testing." *Language Testing*, 2/1, 31-40

Stansfield, C. 1992. "ACTFL Speaking Proficiency Guidelines." *ERIC Digest ED347852*. www.eric.ed.gov.

Teasedale A. & Leung, C. 2000. "Teacher asses and phychometric theory: a case of paradigm crossing?" *Language Testing.* 17/2, 163-184.

Thrasher, R. 2004. "The Role of a Language Testing Code Ethics in the Establishment of a Code of Practice." *Language Assessment Quarterly*. 1/(2&3), 151-160.

UCLES CELS Examination Handbook. No Date. Available at: http://www.cambridgeesol.org/support/dloads/cels/cels_speaking.pdf Accessed on 02.02.2006.

UCLES, 2001. *First Certificate in English Handbook*. Cambridge: UCLES.

Upshur, J. & Turner, C..E. 1999. "Systematic effects in the rating of second-language speaking ability: test method an learner discourse." *Language Testing*. 16/1, 82-111.

Weir, C.J., 1990. *Communicative Language Testing*, Englewood Cliffs, NJ: Prentice-Hall Regent.

Weir, C. 2003. "A survey of the history of the Certificate of Proficiency in English (CPE) in the twentieth century." In Milanovic, M. & Weir, C. Eds. *Balancing Continuity and Innovation. A History of the CPE Examination 1913-2002***,** CUP

Weir, C.J. 2005a. *Language Testing and Validation*. Houndmills, Basingstoke: Palgrave Macmillan.

Weir, C.J., 2005b. "Limitations of the Common European Framework for developing comparable examinations and tests." *Language Testing*. 22/3, 281-300.

References

Clark, J.L.D. 1972. *Foreign Language Testing: Theory and Practice.* Philadelphia PA: The Center for Curriculum Development

Messick, S. 1989. "Validity." In Linn, R.L. *Educational Measurement*. New York: Macmillian/American Council on Education, 13-103.

Shohamy, E. 1997. "Testing methods, testing consequences: Are they ethical? Are they fair?" *Language Testing*, 14, 340–349.